

Assessment of Japan's universities long overdue

The relationship between Japan's universities and the education ministry is too undemanding to allow critical appraisal of research. Rigorous external evaluation must become the norm.

Assessment and appraisal are still alien concepts in Japan, particularly in the world of academia, where respect for the professor is paramount. Thus, it is not surprising that resistance from university faculty members to the introduction of external research assessment is widespread (see page 378).

But the writing is on the wall. Japan's student population is plummeting, owing to a long-standing decline in the birth rate, and universities are having to compete to survive. An increasingly visible part of that competition is on the research front where ever more sophisticated bibliographic analysis by third parties will reveal the performance of university departments whether they like it or not. Furthermore, while the Japanese government continues to be comparatively generous in funding science in a time of economic recession, there are growing calls in government circles for science to be targeted towards socio-economic needs.

That is a dangerous trend and Japanese academics should not bury their heads in the sand and hope that it will go away. Rather, they should face the inevitable and take the initiative in establishing a nationwide system of research assessment that takes account of the need for academic freedom. There are plenty of models around to learn from and, close to home, Singapore and Hong Kong are in this respect much more advanced (see *Nature* 389, 113–117; 1997).

External reviews of university departments and institutes have been carried out on an ad hoc and sporadic basis since early 1993 when Akito Arima, current minister of education, sent shock waves through Tokyo University by bringing in a team of overseas

researchers to assess its physics department. But such reviews are few and far between. They follow no consistent pattern. There is no system of benchmarking the assessments for comparison with other institutions, and the results are seldom made public. Furthermore, in only very rare cases have significant changes been made as a result of the reviews. Their prime motivation often seems to be to gain support for some new pet project, rather than being a genuine attempt to assess both the strengths and weaknesses of institutions.

The time is ripe for change. Arima, Japan's leading advocate of external review, has been appointed as head of both the education ministry, which oversees universities, and the Science and Technology Agency — both to be merged by 2001. Agency officials seem to have a greater awareness of the need for external assessment than their counterparts in the education ministry who, until now, have been only too willing to accept the sop of 'self-evaluation' proffered by powerful university professors and adopted by most universities in Japan as the norm.

In the long run, government officials will need to concentrate on value for money, and Japan will find its science increasingly evaluated on the international arena by complete strangers. This *gaiatsu* ('external pressure') will, along with the power of a handful of reformers in Japan, bring about the change that is inevitable. The academic community needs to work constructively with this change rather than seek to ignore it. This is essential if Japan is also to tackle the even more complex issue of giving much needed autonomy to universities to run their own affairs. □

Hot seat in the kitchen

Last week's announcement that Britain will set up a Food Standards Agency opens a door for someone special.

Applications will be invited later this year for the post of chief executive of a new agency, responsible for food safety — researching, enforcing, advising government and informing the increasingly worried British public.

If your application is successful, you will have the satisfaction of (not before time) bringing the United Kingdom into step with other countries in detaching the regulation of food safety from conflicting agricultural interests and answering to a health ministry. You will control funds transferred from other ministries, plus up to £50 million (US\$82 million) levied from the food industry. You will have a powerful hold on the food safety agenda, in contrast, for example, to the fragmented situation in the United States. Because many of the issues are international in scope, you will have the opportunity to set an example to the world in assessing risks to health and, in collaboration with other ministries, the environment. The scientific research you control will be exemplary if (hopefully) unexciting in its attendance to public concerns about, for example, allergenicity from genetic modification. Your power to raid agricultural or other premises where standards are suspected to be

lax should give rise to occasional dramas.

You can set new international standards in the access you provide to the public to basic and not so basic facts about food and some of the controversial technologies used in its production — communicating, for example, some sense of perspective of the comparative risks of contracting *Salmonella* poisoning from uncooked poultry (relatively significant) and digesting genes from genetically modified crops (negligible based on all the evidence so far).

You will need a cool head as you find yourself countering industrial lobbyists in insisting on more research before the introduction of novel foods and as you enrage some green groups with your judgement that some highly unnatural crops have indeed been adequately tested. Your minister can invoke national security and shut you up — a power that may at times need to be resisted. And as the person ultimately responsible for several scientific advisory committees, you should be prepared for sleepless nights as equivocal evidence emerges that, if made publicly available, might bring some part or other of the food industry to its knees.

Still interested? □



Curb on foreign computers puts damper on US climate modelling

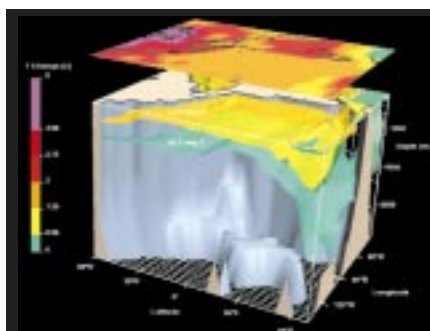
[WASHINGTON] The United States has fallen behind other countries in its ability to model long-term climate change, says a panel of the National Research Council (NRC), the executive arm of the National Academy of Sciences.

The lag has been partly caused by US researchers being prevented from buying powerful supercomputers from Japan, says the NRC climate research committee's report. It adds that the US climate modelling community is likely to "remain behind the rest of the world in terms of computational facilities for the next several years".

Although US researchers still hold the lead in small- and intermediate-scale models, some feel that they lag behind in developing higher-resolution, integrated climate models of the kind needed to address policy issues raised by the Kyoto climate treaty.

"The reliance of the United States upon other countries for high-end climate modelling must be redressed," the panel argues, for reasons that have more to do with policy than with science. It points out that decisions on the Kyoto Treaty could have far-reaching economic impact, and US decision-makers should not depend on simulations done by countries "with different priorities than those of the United States".

The study was led by Thomas Karl of the National Climatic Data Center in Asheville, North Carolina. It was prompted by a letter written in 1995 by several prominent climate modellers to leaders of the US Global Change Research Program (USGCRP). The scientists had warned of a crisis due to a lack of computing power, and said the United States had already been eclipsed by Germany and the United Kingdom as world leaders in climate modelling, with Japan gaining rapidly.



Hot topic: powerful computers are needed to model the impact of CO₂ doubling.

The panel based its conclusions in part on an analysis carried out by Bill Buzbee, director of scientific computing for the National Center for Atmospheric Research (NCAR) in Boulder, Colorado. NCAR's most powerful computers can sustain just 5 gigaflops (5 billion operations per second) when running a single application, whereas comparable centres in Australia, Canada, England and elsewhere have systems that can sustain from 20 to 100 gigaflops, said Buzbee.

Much of the blame goes to a controversial Commerce Department 'anti-dumping' ruling of 1997, which blocked the Japanese supercomputer maker NEC from selling powerful SX-4 machines to NCAR (see *Nature* 389, 432; 1997).

The decision "seems to have discouraged other institutions within the United States from considering the purchase of foreign computers," says the NRC report. "This lack of routine access by [US] climate researchers to the world's most powerful computers has become a quite serious problem."

Jerry Mahlman, who heads the Geophysi-

cal Fluid Dynamics Laboratory (GFDL) in Princeton, New Jersey, a leading US climate modelling centre, was one of the outside reviewers of the NRC report. Even though Japanese machines offer more, he says, his laboratory, which wants to advance its computing capability to 150 gigaflops, assumes it will have to buy American computers.

The NRC committee found that US climate modellers face other problems, including the lack of an integrated government strategy for tackling important scientific and policy questions. There is little standardization of key computer subroutines or model outputs, for example. Component models of the atmosphere, oceans and land are not always incorporated effectively into more comprehensive models.

Adding to the problem, it says, is that US researchers do not always have "full, open and timely access" to output from foreign climate models. The situation could get worse as commercial considerations — for example, the privatization of European meteorological data — come into play. But it warns that any proposed solution to this non-reciprocity "should not involve the imposition of access restrictions to US data".

The panel's conclusions have already generated debate. Mahlman believes it may overstate the gap between US and other climate modellers. His own centre is running simulations at 15 gigaflops, he says, and publishing results similar to those from foreign modelling groups. He also worries that the NRC recommendations could be taken as jingoistic. "These so-called foreigners are our friends and colleagues," he says, and collaboration is common.

US supercomputing capability may not trail the competition for long. The Department of Energy is expected to receive \$70 million of a proposed \$366 million increase for information technology (see *Nature* 397, 285; 1999) which aims to increase computing power by three orders of magnitude to 40 teraflops (40 trillion operations per second).

The first scientific applications for the Department of Energy's advanced supercomputing capacity would be combustion systems and research on "global systems". The computers would come on-line in 2003, and support the Intergovernmental Panel on Climate Change's fourth assessment in 2005.

But Mahlman echoes one of the panel's cautions: fast machines alone won't improve climate prediction. There must be investment in the underlying science, he says, to create a "balance between supercomputing power and brain power".

Tony Reichhardt

Geophysicists call for action on global warming

[WASHINGTON] The American Geophysical Union (AGU), one of the leading scientific organizations in the United States, came out last week after months of internal debate in favour of action to counter global warming.

In a cautious statement, the council of the union recommended "the development and evaluation of strategies such as emissions reduction, carbon sequestration, and adaptation to the impacts of

climate change". The present level of scientific uncertainty, said the AGU, "does not justify inaction in the mitigation of human-induced climate change".

The 600-word statement was based on an extensive review of the peer-reviewed literature but no new scientific evidence. Because uncertainty about the causes and effects of climate change "will never be completely eliminated... science cannot be the sole

source of guidance on how society should respond to climate issues", it said.

Vice-president Al Gore immediately focused on two of the AGU's points: that there is a "compelling basis for legitimate public concern" about climate change, and that the lack of certainty does not justify inaction.

But the conservative Competitive Enterprise Institute claimed that the AGU had "succumbed to political correctness".

T.R.

Rubbia proposes a speedier voyage to Mars and back

[MUNICH] The Italian Space Agency has carried out a feasibility study on a new type of propulsion engine that would allow interplanetary travel at speeds fast enough to make manned trips to Mars a realistic proposition. The agency is planning to investigate the idea further, says its science director, Giovanni Bignami.

The idea was developed by Italian Nobel prize winner Carlo Rubbia, the former director of the European Laboratory for Particle Physics (CERN) in Geneva, now professor of physics at the University of Pavia, near Milan. His proposed engine uses fission fragments of the element americium to heat a propulsion gas directly.

An interplanetary propulsion system called NERVA is being studied by the US space agency NASA. NERVA is based on a classical nuclear reactor, using solid fuel rods to transfer energy to a propulsion gas. The amount of energy generated is therefore limited by the melting point of the fuel rods.

In contrast, Rubbia's engine has the fuel, an isotope of americium, painted thinly on the inside wall of a chamber containing hydrogen gas. The chamber is surrounded by graphite, which acts as a reflector for neutrons bombarding the americium fuel. Neutrons would be provided both by a natural source and by the americium fission process itself.

Reflected to and fro across the chamber, the neutrons pass frequently through the americium, and have a high probability of colliding with americium atoms. The fission products of such collisions scatter into the chamber and deposit their energy in the gas.

According to Rubbia's predictions, the gas could be heated in this way to up to 14,000 K, and could be used to generate speeds of up to 40 kilometres per second. A manned trip to Mars and back could be achieved in around a year using such a propulsion system. In contrast, a trip using NERVA is calculated to take four years.

"Survival of humans in space is intrinsically limited by cosmic radiation, even on the surface of Mars with its thin atmosphere and weak magnetic fields," says Bignami. "So cutting mission duration by several years in itself makes the difference between a dream and a feasible mission."

Rubbia stresses that the concept is "just an idea" but is "based on easy scientific principles and the development of a machine which would theoretically be very reliable because it would have few moving parts". The Italian agency is now planning a design phase involving industry and international partners, says Bignami.

Alison Abbott

AAAS head wants to see scientists on school boards

[ANAHEIM, CALIFORNIA] US scientists were urged last week to actively promote science education by M. R. C. Greenwood, chancellor of the University of California at Santa Cruz and president of the American Association for the Advancement of Science (AAAS).

Speaking at the association's annual meeting, Greenwood said scientists should re-examine their responsibility for meeting the needs of society. She called for the AAAS to launch a campaign to ensure there is at least one scientist, engineer or scientifically literate person on every school board.

Greenwood argued that the falling academic performance of US schoolchildren threatened not only the next generation of scientists, but the whole national support system for science. She pointed out that a 1995 study found that US 12th graders scored nearly bottom of 50 nations tested in science skills, and that even advanced placement students in subjects such as physics ranked extremely low.

Greenwood placed the blame for this on poor textbooks, on teachers with little experience of studying science, on a lack of attention to women and minorities, and on a vocal minority of parents and community leaders who are hostile to science education. These factors have allowed an aversion to science to



Greenwood: plans to tackle school science.

seep into the classroom, said Greenwood.

She urged scientists to take a lead from grassroots groups who target school boards to further their political agendas. Religious fundamentalists, for example, have used membership of school boards to prohibit the teaching of evolution or to require announcements that

evolution is merely a theory.

Although the AAAS would not run political campaigns itself, she said it should encourage scientists to take public office. Such scientist-activists would seek to convince school boards that science, engineering and mathematics are critical life-skills for all students. The AAAS would provide materials and a network of experts to support members' efforts to improve pedagogy, curricula, textbooks and student assessment.

Greenwood, who presented the plan to AAAS membership at the meeting, plans to develop a proposal for the board and seek funding for a workshop on her proposed strategy by the middle of 1999. **Sally Lehrman**

Harvard reveals plans for research centres

[BOSTON] Harvard University announced last week that it will invest between \$150 million and \$200 million over the next five years in new programmes in science, with a special emphasis on interdisciplinary activities.

A panel of faculty members, set up last summer to identify the most promising research directions, has recommended the establishment of several new research centres. Two of these — the Center for Genomics and Proteomics and the Center for Imaging and Mesoscale Structures — have already been approved for funding.

Three more centres are being considered: a neuroscience facility to explore the connections between single nerve cells and an organism's behaviour by studying large groups of neurons simultaneously; a content engineering facility to develop Web search engines to mine scientific databases; and a programme to examine links between global climate change and biological evolution.

The new programmes, says chemist Jeremy Knowles, dean of the university's arts and sciences faculty, will cross departmental boundaries and foster research and education along interdisciplinary lines.

"The old paradigm of a single investigator applying for grants and then assembling a lab and a research group has become less viable in a number of fields," he adds.

Investigators at the genomics centre will study genes and their proteins, exploring animal behaviour, evolution, the genetic causes of disease and other complex biological phenomena. Those at the imaging centre will mainly study extremely small structures on the scale of 1 to 100 nanometres, with applications including the design of smaller computer processing and fibre-optic communication systems, the creation of biomaterials and drugs, and the development of optical and scanning technologies for medical diagnostics.

Harvard is eager to get the first two centres operating, says Knowles. Although some construction is planned, they will initially use existing facilities, he says.

About \$70 million dollars has been allocated for equipment, staff and construction projects. Knowles says other centres may be created as Harvard strives to pursue "the most exciting and fruitful areas of inquiry". **Steve Nadis**

Treaty nations 'are exploiting Antarctica'

[SYDNEY] Forty years after the Antarctic Treaty was signed, the first meeting of ministers on the frozen continent generated an urgent sense last week that measures to protect the territory are failing.

Simon Upton and Robert Hill, environment ministers of New Zealand and Australia, are pressing for action against the lucrative but illegal fishing industry in the Southern Ocean. They want to see at least A\$600 million (US\$378 million) a year spent to combat it. Along with Chile, Argentina, France, Norway and the United Kingdom, the two countries have claims over sectors of Antarctica.

Last week's meeting, held under the auspices of New Zealand with support from the United States and Italy, attracted senior representatives (including 13 ministers) from 24 of the 43 nation members of the Treaty to New Zealand's Scott Base.

At the meeting, Upton characterized the largely covert fishing industry as "a threat to the credibility and integrity" of the treaty and its spin-off, the 27-member Convention on the Conservation of Antarctic Marine Living Resources (CCAMLR).

"The great bulk of the uncontrolled fishing for the [Patagonian] toothfish in the waters managed by CCAMLR has been, and is being, carried out by companies and citizens from parties either to the Antarctic Treaty or the CCAMLR — an awful indictment of us," said Upton.

Hill revealed estimates by Australian sci-



Toothsome: the current level of poaching could wipe out toothfish in three years.

entists that, at the present rate of poaching, the toothfish will be extinct in only two to three years. The long-line fishing method also endangers populations of other fish species and albatross.

Representatives formalized an agreement with New Zealand for scientific collaboration. Tony Press, director of the Australian Antarctic Division, and Gillian Wratt, chief executive of Antarctica NZ, say they plan joint studies of fish and krill populations to support monitoring and conservation.

Australia spends A\$60 million annually

on Antarctica while New Zealand spends about NZ\$17 (US\$9) million. Most of this money is spent on maintaining bases on the continent and Sub-Antarctic islands. Additional funds for research come from competitive grants to university staff. Both countries admit that their small navies and air forces can mount only inadequate surveillance.

Logistic sharing is likely later, especially over transport. New Zealand can fly to its base from Christchurch, while Australia has become increasingly hampered by dependence on a chartered icebreaker (see *Nature* 390, 109; 1997). The *Aurora Australis* has suffered three breakdowns since last July, forcing Australia to acknowledge last week that its Antarctic science programme is "in disarray". A Norwegian ship has been chartered for the voyage cancelled last month, but it has to leave from Cape Town.

Although last week's meeting carried no formal force, some ministers used it as a springboard for more concerted lobbying before an April meeting of the CCAMLR in Brussels, and a full meeting of Treaty ministers in the Peruvian capital, Lima, in June.

Measures proposed to save the toothfish are based on introducing mandatory systems for monitoring vessels by satellites; placing scientific observers on board licensed vessels to report on other vessels sighted; uniform licensing and marking systems for vessels; and making it difficult to land illegally caught fish and sell it in the main markets of Japan and the United States. **Peter Pockley**

German science council clashes with anti-nuclear government

[MUNICH] In a report that appears to conflict with the new government's antipathy to nuclear power, the Wissenschaftsrat, Germany's influential science council, has called for an across-the-board 30 per cent increase in the energy research budget over the next three years.

The council says funding needs to be more efficiently coordinated between the federal government and the 16 *Länder* (states), criticizing universities for the fragmentation of their energy research.

But it also recommends that the breadth of research topics, including nuclear fission and fusion, should be maintained, stressing the "need to maintain nuclear competence".

Germany's Social Democrat and Green coalition government hopes to phase out nuclear energy. The Greens have criticized the Wissenschaftsrat's demand that all energy options should be left open, as well as its refusal to compare renewable and non-renewable energy sources. "The opportunity to reassess and reorganize energy policy from a scientific point of view has been

wasted," says Hans-Josef Fell, the Green parliamentary spokesman on research.

A Wissenschaftsrat committee, advised by industrial experts, spent almost two years reviewing energy research in Germany. It warns that, without more money for research, Germany risks losing its international competitiveness in areas such as power technology, reactor safety and waste disposal technology.

Germany spends only around DM1.1 billion (US\$650 million) per year — 0.015 per cent of its gross domestic product — on energy research. Japan spends six times more, France three times more and the United States twice as much.

The report recommends additional funding for interdisciplinary research on the efficiency of fossil energy sources, which supply 87 per cent of the primary energy consumed in Germany, but contribute heavily to greenhouse gases.

It also recommends greater investment in renewable energy sources, such as solar radiation, biomass, water, wind and

geological heat, which provide around two per cent of Germany's primary energy.

Nuclear research, it says, should focus on reactor safety and nuclear waste disposal. But the need for international cooperation in reactor safety means that existing reactor technology should be further developed. It urges that all aspects of nuclear physics and radiochemistry should be taught at universities, warning that Germany could fall behind because of recent spending cuts.

The Greens agree with the proposal for more research in reactor safety and waste disposal, but oppose additional nuclear research. In contrast, the Social Democrat's science advisers have indicated that phasing out nuclear energy should not necessarily end basic nuclear research.

The Greens and the science council agree that the public and social scientists should be given more influence in decision-making on energy issues. The Greens also agree with the Wissenschaftsrat's proposal to encourage a broad public debate about the risks of fusion technology. **Quirin Schiermeier**

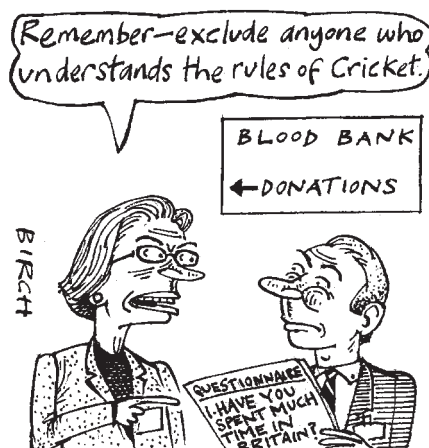
Fear of BSE risks could hit US blood banks

[WASHINGTON] In a bid to reduce the theoretical risk of introducing new variant Creutzfeldt-Jakob disease (vCJD) into the United States, advisers to the US Food and Drug Administration (FDA) are considering whether to recommend that people who have spent time in Britain since 1980 should be barred from donating blood.

The FDA's Transmissible Spongiform Encephalopathies Advisory Committee voted in December by nine to six to consider recommending that the agency adopt some kind of restriction on such donors. But it did not specify how long potential donors would have had to have spent in Britain to be excluded, saying it needed more data.

Any such move, however, is likely to meet opposition from those responsible for blood banks, which already have problems obtaining adequate supplies. According to an American Red Cross official, excluding all donors who visited or lived in Great Britain between 1984 and 1990 would reduce the US blood supply by 10.7 per cent, and more than one million new donors would be required to replace those excluded.

The years 1984-90 are considered the peak risk period for contracting vCJD through eating meat contaminated with the infectious agent that causes bovine spongiform encephalopathy ('mad cow disease'). The committee will debate the matter further and may make a formal recommendation in June. Before doing so, it will consider



data analysing how long typical US blood donors spent in Britain in the past 19 years.

No case of vCJD has ever been attributed to blood or a blood product since the new variant disease was identified. But most committee members felt that caution was needed given that the extent of the disease in Britain is unknown, and transmission by blood remains a theoretical possibility.

So far, 35 'definite and probable' cases of vCJD have been established in Britain. If in five years Britain has 1,000 cases or more, said Paul Brown, a senior research scientist in the Laboratory of Central Nervous System Studies at the National Institute of Neurological Disorders and Stroke, who chairs the advisory committee, then adopting restrictions would be seen as "a very shrewd posture".

But representatives of blood banks at the meeting challenged this view. "The national blood supply is teetering on the brink of inadequacy," said Merlin Sayers, director of the blood bank at the Carter Blood Center in Bedford, Texas. "For us to entertain even a five per cent additional donor [loss] to prevent something which has not happened is going to be a national experiment doomed to failure, and it will jeopardize patient care."

Blood-bank officials also argued that the loss of donors would require recruiting new donors who present a small but real — rather than theoretical — risk of introducing hepatitis C or HIV into the blood supply.

Jay Epstein, the director of the FDA's Office of Blood Research and Review, conceded that it was hard to balance risks. "The problem for the FDA is that we have to factor into our thinking all the possible consequences of a change in policy," he said.

Some committee members hope that a reasonable compromise can be agreed on if the new data indicate that most travellers to Britain spent only a brief time there. A policy excluding only those who made lengthy visits would have far less impact on the US donor pool, they suggested.

"If you focus on the highest-risk people, you will get almost all of the protection and do very little to limit the supply of needed blood," says Peter Lurie, a medical researcher with the Public Citizen Health Research Group in Washington DC. **Meredith Wadman**

Asian scientists call on Unesco to protect indigenous knowledge

[BANGALORE] A group of scientists and social scientists from eight Asian nations last week called for the United Nations Educational, Scientific and Cultural Organization (Unesco) to set up a global fund to conserve and promote 'indigenous and civilizational knowledge systems' (ICKS).

The proposal was made at an international symposium held in Bangalore, India, as part of the preparations for the World Conference on Science in the Hungarian capital, Budapest, in the summer (see *Nature* 396, 299; 1998). It was sponsored jointly by Unesco and the Indian government, to focus on the social dimensions of science and technology.

Those present urged the UN body to organize two international conventions. One would focus on sustainable consumption, aimed at regulating the use of fast-eroding natural resources. The other would address the brain drain, to compensate developing nations for their loss of trained scientists and technicians to developed countries.

Researchers from China, Iran, Uzbekistan, France, the USA, Bangladesh,

Nepal, Sri Lanka and Thailand, as well as India, attended the meeting.

Officials of the National Institute of Advanced Studies (NIAS) in Bangalore, which organized the three-day meeting, said the recommendations form part of a nine-point action plan. Delegates hope to see this included in the declaration *Science agenda — a framework for action*, which is expected to be adopted at Budapest.

As well as establishing a global fund for promoting ICKS, the proposals call on Unesco to design a series of projects that recognize the economic value of indigenous technologies and study how best to protect them, "leading to an international convention on protection of intellectual property rights in ICKS".

Delegates also suggested that Unesco draw up an international code of ethics for science and technology that would address a wide range of issues. These include the regulation of 'knowledge monopolies' — companies or groups of companies exercising monopoly power over strategically important knowledge.

Other proposals were that Unesco should support training programmes for scientists on legal and policy issues, and for an international conference to examine the future impact of culture on science.

All these proposals will be put to the Budapest conference, together with the Bangalore communiqué: a document drawn up primarily by the Indian government which describes how society's expectations of science might be met in the 21st century.

The Bangalore symposium adopted an amended version of Unesco's declaration on science and the use of scientific knowledge, reflecting India's concerns on traditional knowledge. (See <http://helix.nature.com/wcs> for a fuller report.)

In his opening speech the Indian science minister, Murli Manohar Joshi, expressed concern that information technology could create new inequalities. He urged nations to ensure that, while newly produced scientific knowledge is legally protected, "those who created and preserved various forms of traditional knowledge so faithfully over centuries are not exploited". **K S Jayaraman**

US budget 'freezes' spending on research

[WASHINGTON] Hopes of substantial increases in science spending in the United States suffered a major setback this week. President Bill Clinton's proposed budget for fiscal year 2000 would essentially freeze spending on research and development in the largest science agencies at 1999 levels.

Clinton proposed a 2 per cent increase at the National Institutes of Health (NIH), the lowest for this agency for many years, and a reflection of what the administration — and some outside — consider to be the excessive 15 per cent boost that NIH received last year.

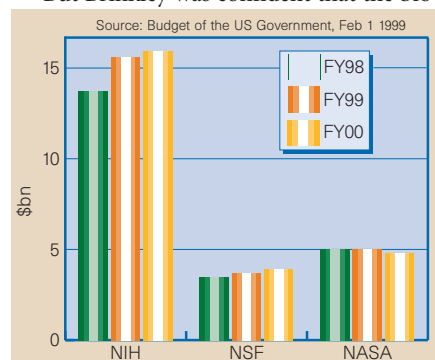
The National Science Foundation (NSF), which supports most non-biomedical research at US universities, did better than other agencies, with an increase in its research budget of almost 7 per cent. But more than half of that will support a new, multi-agency initiative in information technology, leaving most research directorates at NSF with increases below 3 per cent.

The outlook was even bleaker at other agencies that fund university research, such as the Department of Defense, which plans to reduce its research budget by 4 per cent.

These cuts, and additional reductions in military development and testing, mean that the total amount of money for research and development in Clinton's 2000 budget would decline slightly, from \$79.2 billion to \$78.2 billion. However, the proposal would meet the administration's long-standing objective that civilian and military research and development should be funded equally. When Clinton came to power in 1993, military work consumed 60 per cent of the total.

Research lobbyists reacted with disappointment to an overall budget less auspicious than the one hinted at by extensive leaks preceding Monday's announcement. "We're disappointed that the president chose not to continue the momentum he helped to create last year," says William Brinkley, vice-president of the Baylor College of Medicine in Houston, Texas, and current president of the Federation of American Scientists for Experimental Biology. "This sends a terrible message to young investigators."

But Brinkley was confident that the bio-



Proposed R&D spending at top science agencies.

Highlights of the budget request for 2000

The budget of the **National Institutes of Health (NIH)** will increase by \$320 million, or 2.1 per cent, to \$15.9 billion. The major NIH institutes will each receive increases at around that level — just enough to cover inflation. NIH says the proposal would allow it to support 7,617 new investigative grants in the fiscal year 2000, compared with the 9,171 grants that it will issue this year.

The National Science Foundation fares best of the large agencies, with a proposed increase of \$250 million, or 6.9 per cent, taking its budget close to \$4 billion. As well as the new computing initiative, costing \$146 million, the agency plans a new \$50 million programme in biocomplexity — a favourite subject of Rita Colwell, its new director.

The Department of Energy's \$2.8 billion request for science is \$138 million more than last year. The largest increases go to the new supercomputing initiative (\$70 million) and to advance construction of the Spallation Neutron Source facility at the

Oak Ridge National Laboratory in Tennessee (\$214 million). High-energy physics (\$697 million), nuclear physics (\$343 million) and fusion (\$223 million) remain essentially flat.

NASA's budget dips less than one per cent to \$13.58 billion, with space science (\$2.2 billion), life and microgravity science (\$256 million) and Earth science (\$1.46 billion) all hovering around last year's levels. New science initiatives include a communications relay for Mars-orbiting spacecraft and a low-cost Mars 'airplane' to scout sample return sites.

The Department of Defense, which remains the most important sponsor of US university research in computer science and engineering, will freeze its investment in basic research at \$1.1 billion, and sharply cut applied research by \$200 million, to less than \$3 billion. The Pentagon's expenditure on development and testing will also fall sharply, despite generous increases in total military spending.

The National Institute of Science and Technology is seeking to raise the budget of its Advanced Technology Program by \$40 million, to \$240 million, and wants over \$100 million for new laboratories, mainly the Advanced Measurement Laboratory, under construction at Gaithersburg, Maryland.

The \$535 million request for research and development at **the Environmental Protection Agency** is about the same as last year's budget, after factoring out congressional 'earmarks'. Small increases go to climate change research, investigating the health effects of particulate matter, coastal environmental monitoring, and assessment of human health risks.

The budget of **the US Geological Survey** would climb 5 per cent to \$838.5 million, with extra funds for a disaster information network, amphibian research and monitoring, and public use of geophysical data.

medical research lobby can still persuade Congress to grant another 15 per cent increase this year for NIH. Neal Lane, Clinton's science advisor, defended the 2 per cent increase at NIH, pointing out that it would take the agency's 2000 budget to the level proposed by the administration a year ago, before Congress granted this year's increase.

The budget proposal for the space agency NASA contains few surprises, but enough money to maintain existing science projects. Dan Goldin, NASA's administrator, once again touted his agency as a model of government 'reinvention', saying: "For the sixth year in a row, NASA's budget has declined while productivity improves."

The most important fresh research initiative was the \$366 million, multi-agency information technology scheme announced last month by vice-president Al Gore (see *Nature* 397, 285; 1999). Rita Colwell, director of the NSF, which will lead the initiative,

pledged that all disciplines that use super-computer modelling would be boosted. "All of science is going to benefit from this, and funds will go to the best proposals that come in," she said.

The administration requests \$1.8 billion for the second year of its Climate Change Technology Initiative, an increase of one-third. The total includes \$400 million of tax credits. Most of the money would go to the Department of Energy, to develop renewable and fossil-fuel energy sources and promote energy-efficient technology. But \$32 million of it is for scientific research, including work on microbes that could sequester carbon to help meet greenhouse-gas reduction targets.

The Environmental Protection Agency's share in the climate initiative rises from \$109 million last year to \$216 million, most going to develop energy-efficient technologies and for tax incentives for decreasing energy consumption. **Colin Macilwain & Tony Reichhardt**

Leaked CNRS report raises fears and hopes on job security

[PARIS] A leaked report has heightened tensions between researchers at the French Centre National de la Recherche Scientifique (CNRS) and the ministry of national education, research and technology.

The draft assessment of CNRS was leaked to the newspaper *Le Monde* last week. It was written by the chair of an external visiting committee, the eminent French biologist Pierre Chambon; other members of the panel include David Baltimore, president of the California Institute of Technology, and Sir Ronald Oxburgh, rector of London's Imperial College of Science, Technology and Medicine.

Its contents have been reported as questioning the lifelong job security enjoyed by CNRS researchers. Many scientists are already suspicious that planned reforms by the ministry will result in the agency being dismantled (see *Nature* 396, 607; 1998).

Chambon has been quick to play down the conclusions of the report, arguing that it is only a "working document" that contained no radical new insights beyond the widely acknowledged lack of mobility in the French research system and the difficulties faced by young researchers in setting up independent

laboratories in a system dominated by established scientists.

"No member of the committee recommended dispensing with the civil-servant status of researchers," Chambon emphasizes. He argues that, at most, the report supports the widely floated idea of creating a single corps of researcher/lecturers

whose time allocated to research and teaching duties would vary during their career according to their productivity.

Meanwhile, several members of the committee are said to be annoyed that Chambon circulated a draft they had not approved. Although members commented on a first draft, they had not seen the circulated version taking these comments into account. Chambon says the new draft was not intended to be published or made public, but was sent to CNRS and the ministry to test whether it was likely to be acceptable.

Oxburgh, who confirms that he had not seen the draft, says, "CNRS is an organization with a tremendous tradition and has produced some outstanding science. They have the problems inherent in any organization that appoints people for life for the sole purpose of doing research." **Declan Butler**



Chambon: draft was 'a working document'.

Japanese labs balk at bid to boost external evaluation

[TOKYO] Government-led efforts to promote the external evaluation of research in Japanese government institutes and national universities are meeting resistance from the laboratories themselves.

The first report on Japan's research evaluation system concludes that only a third of evaluations at research institutes run by science-related ministries are carried out by external assessors. The Science and Technology Agency (STA) report, released last week, says external evaluations are also uncommon in universities.

The need for greater evaluation was emphasized in guidelines drawn up in 1996 as part of the Science and Technology Basic Plan, which aims to double Japan's science spending by 2001.

The guidelines, which also apply to individual projects, say that to be effective, they should be applied by an external review system. Although voluntary, they give a framework by suggesting timing, method and objectives. They also call for the results of evaluations to be made public, and for funding to be based on these.

An STA spokesman says that, as well as assessing the quality of research, evaluations are necessary to determine whether research is meeting social and economic needs, such as the creation of new industries. "It is important to find the priority areas for limited government funds, and to gain public support for our large investment in science and technology," he says.

The guidelines do not apply to universities, on the grounds that the independence of research and the balance between education and research must be respected. A separate 1996 proposal from the Council on University Education — an advisory body to the Ministry of Education, Science, Sports and Culture (Monbusho) — emphasized self-evaluation and in-house assessment of university research, but limits external evaluation to large projects.

According to the report, all but one of Japan's 98 national universities, and 381 out of all 587 Japanese universities are carrying out some form of research evaluation. But only a handful, such as Tokyo University's Research Centre for Advanced Science and Technology (RCAST), use external reviewers, and most of the results are confidential.

This contrasts with the National Research Institutes for Joint University Use, which provide facilities and equipment for domestic and international researchers, such as the High Energy Accelerator Research Organization in Tsukuba and the National Institute for Genetics (NIG). Eighty per cent of these

have introduced external evaluation and have made the results public.

"University researchers are becoming increasingly aware of the importance of research evaluation, and many see it as a way of demonstrating the quality and accountability of their research," says Yoshiki Hotta, director of NIG and formerly director of Tokyo University's Genetics Research Laboratory. "But it may still take some time for the universities to fully adapt to the system."

Hotta emphasizes the importance of evaluation in improving prospects for young researchers. He plans to use evaluation at NIG to get rid of incompetent researchers and to introduce postdoctoral researchers.

Many argue that the lifetime tenure system that plagues Japanese universities with bureaucratic and conservative values can only be eradicated by external reviews. The system has long been criticized for inhibiting competitiveness at Japanese universities.

Some researchers say the main difficulty is the lack of a single framework. "External evaluations are carried out independently by departments and laboratories within the university, so the criteria of evaluations differ from one research group to another," says Etsuo Niki, director of RCAST.

Universities are under pressure to reform following a proposal to turn all 98 national universities into semi-autonomous institutions (see *Nature* 395, 730; 1998). Although strong resistance from universities and Monbusho meant that the proposal was omitted from the final version of the government's 1997 administrative reform plan, it resurfaced after last year's appointment of prime minister Keizo Obuchi.

Despite Obuchi's determination to reorganize Japan's administration, the plan continues to meet opposition from leading academics, such as Leo Esaki, the Nobel prizewinner and former president of Tsukuba University, who has claimed that it "will only make things chaotic".

But the strongest opposition comes from Akito Arima, the education minister and recently appointed director-general of STA, who has declared that such changes will be detrimental to research and education, as universities do not have the capacity to manage their increased autonomy.

In a bid to resist the government's plans on autonomy, the Council on University Education, chaired by Arima, plans to set up an independent body to assess national universities. Ironically, as president of Tokyo University in 1993, Arima introduced external review to the physics department (see *Nature* 362, 387; 1993). **Asako Saegusa**

Tougher EU copyright rules come under fire

Declan Butler

The European Science Foundation warned last week that a proposed European Union directive, intended to ensure that copyright owners receive full legal protection when their work is distributed in digital form, could weaken the 'fair use' arrangements enjoyed by scientists.

[PARIS] A directive proposing changes to European copyright laws will be debated this week by the European Parliament. But the European Science Foundation (ESF) is not alone in claiming that, if left unchanged, this could harm the competitiveness of European research.

The "Directive on the harmonisation of certain aspects of copyright and related rights in the information society" (<http://europa.eu.int/comm/dg15/en/intprop/intprop/copyen.pdf>) was proposed in December 1997 by the European Commission. It points out that, whereas photocopying the work of information producers had limited economic effects, the ease of producing high-quality copying in the digital environment requires a reassessment of the balance between providing incentives to the creation of original work and facilitating the dissemination of such work to users.

"Exceptions and limitations must be construed in a more narrow way... in order to prevent economic damage to the market of protected works and other subject matter," the directive reads.

Amendments to the text will be voted on this week. The Council of Ministers of the member states must then reach a common position. The text will go through a second reading in parliament before being adopted by the Council of Ministers and promulgated into national laws.

Fighting for exceptions

But Sir Roger Elliott, Emeritus Professor of Physics at Oxford, and chairman of the Committee on Scientific Information at the International Council of Scientific Unions, warns that the directive could tip the balance excessively towards the interests of publishers.

Elliott complains in particular that the draft lacks "a general exception for private study and research", in spite of the fact that this is a traditional right in international copyright legislation.

One commission official argues, however, that exceptions for scientific and educational purposes already vary widely within the community. The exceptions in the draft directive apply to "illustration for teaching and research... for non-commercial ends".

The ESF, in a statement released last week,



Right to copy? The balance between the interests of publishers and scientists is a delicate one.

says it wants this changed to read "for the sole purpose of scientific research or for illustration for teaching". It also wants to include research in the arts and humanities. An amendment to be voted on this week would change the wording to "for the purposes of education, learning and research".

One senior commission official vigorously contests claims that the directive will give too much power to publishers. "I have seen claims that there are virtually no exceptions foreseen for research; this is sheer nonsense," he says. "Many of the arguments are based on lack of knowledge about this very complex field of law; the acquired rights of scientists are well established and are not in question."

The directive also allows an exception for the reproduction rights of libraries, but gives libraries no exceptions for "communicating or making material available to the public", insisting that they must make protected material available on-line via licensing agreements (see *Nature* 397, 196; 1999). This is already the case under existing legislation in several member states, including France, Germany and Belgium.

But EBLIDA, an umbrella association representing 95,000 libraries throughout Europe (<http://www.kaapeli.fi/~eblida/posharmon.htm>), warns that "by seeking to regulate the market too tightly, the commission runs the real risk of heavily restricting access

to information and information products, which is in no-one's interests." It "fears a nightmare future in which nothing can be looked at, read, used or copied without permission or additional payment".

Under the proposed directive, individual member states would also be free to interpret the research exemptions as they liked. The ESF claims this could result in research being treated differently in member states, whereas the stated aim of the directive is to nurture an internal European market in information by aligning regulations in all member states.

A commission official argues, however, that "We are simply trying to provide more legal consistency and certainty; we cannot provide uniform rules as we are not a federal state."

Elliott and the ESF also contest the restriction of exceptions to non-commercial use, pointing out that the distinction between the commercial and the non-commercial is becoming increasingly blurred in academia. The introduction of an exception based on a 'public good' definition of research is a more desirable solution, argues the ESF.

Striking a deal

A commission official admits that the non-commercial notion may be invalid, and that the commission is open to change. The exceptions are "a bargaining position", he says. "The final directive will look different."

Roberto Barzanti, an Italian socialist and rapporteur of the parliament's motion, also believes that researchers' concerns are exaggerated, and that the exceptions will protect scientific liberties. The directive, he points out, is mainly about wider issues such as ensuring that member states extend copyright protection to cover digitally distributed material, and allowing them to control distribution, for example through encryption.

Copyright in scholarly publishing is different, he says, with publishers reliant on a balance between their rights and those of scientists (see *Nature* 397, 195; 1999).

Barzanti also points out that the overall copyright legislation is only one aspect, and that much of researchers' and institutions' rights are determined by the contracts they settle with publishers.

Indeed, one of the proposed parliamentary amendments to the directive states that libraries and other "public good" bodies should negotiate contracts and licences whose terms promote their goals of disseminating information. Elliott agrees that much can be achieved within specific contracts, but nonetheless argues: "It is a different relationship if you are negotiating with the law behind you, or with the law in your face."

"One of the problems has been that while the US scientific community [has organizations that] can mount action against laws, scientists in Europe have no effective focus for dealing with Brussels," says Elliott. □

Former South African science minister may return to cabinet

[CAPE TOWN] As South Africa's Minister of Arts, Culture, Science and Technology, Lionel Mtshali, leaves the cabinet this week to be installed as premier of the province of Kwazulu-Natal, there is speculation that the current premier of the province, Ben Ngubane, will return to the cabinet to head the ministry, as he did from 1994 to 1996.

Ngubane was moved to Kwazulu-Natal to deal with the delicate political situation in this strife-torn province, the only one controlled by the Inkatha Freedom Party (IFP) led by Mangosuthu Buthelezi.

Under an agreement with the African National Congress (ANC), Buthelezi has the right to appoint two members of the cabinet. But Mtshali's performance as a minister (see *Nature* **396**, 298; 1998) apparently led to dissatisfaction among the ANC leadership.

Japanese telescope to seek edge of universe

[TOKYO] Subaru, one of the world's largest optical infrared telescopes, made its debut last week as Japan's National Astronomical Observatory unveiled images from its 'first

light'. The telescope, located near the summit of Mount Mauna Kea in Hawaii, has the world's largest one-piece reflective mirror, measuring 8.2 metres in diameter.

"We hope to look at the very farthest edge of the universe," said Keiichi Kodaira, director of the observatory. Subaru, the Japanese name for the Pleiades, will begin its full operation in April 2000.

China to set up criminal DNA database network

[LONDON] The Chinese government has announced plans to create a national network of DNA databanks of known criminals. The first phase, which will involve collecting 2,500 DNA samples, is expected to be completed by the end of August. Its cost is estimated at US\$300,000.

Council of Europe votes to halt xenotransplantation

[PARIS] The Council of Europe has called for a legally binding moratorium on clinical trials of xenotransplantation — the transplanting of animal cells, tissues and organs into human beings. The parliamentary assembly of the Council, a political body representing 40 countries, voted unanimously for the proposal (see *Nature* **397**, 281; 1999).

The assembly ruled that a moratorium should stay in place until the risks of man-made pandemics were better assessed, and international guidelines established. It called on the council to work towards an international moratorium.

Surgeons call for more organ donors and funds

[LONDON] Britain's Royal College of Surgeons called last week for an increase in research fellowships in transplant surgery to address the lack of interest among trainee surgeons because of limited research opportunities.

In a review of organ transplantation in England and Wales, the college also called for Britain to double its present number of organ donors to 10 million within a decade, as it does not expect the shortage to be met by xenotransplantation within that period.

A survey by British Gas shows that Britons are uncertain about xenotransplantation. Though 91 per cent of respondents said they would consider having such a transplant, 54 per cent admitted that it worried them.

Scientists unite to speed up genome sequencing

[WASHINGTON] The genome sequencing company Celera Genomics has formally

joined forces with researchers at the University of California, Berkeley, to speed up the sequencing of the *Drosophila* genome, according to an announcement last week by the National Human Genome Research Institute (NHGRI).

The joint effort by the company, based in Rockville, Maryland, and Gerald Rubin, an investigator at Berkeley who is funded by NHGRI, will allow the completion of the 150-million-base-pair fruitfly genome this year, according to an agreement signed by Rubin and J. Craig Venter, president of Celera.

The collaboration will provide “an important pilot” for similar public-private cooperation on the human genome, says Francis Collins, the director of NHGRI.

Chemical companies to fund health research

[WASHINGTON] The US Chemical Manufacturers Association (CMA) last week launched a \$100 million, five-year programme to fund basic research into the health and environmental effects of common chemicals. The initiative is part of a \$1.2 billion, six-year effort by the industry that includes toxicity testing of 3,000 chemicals made in large quantities, and research on 15,000 others to see whether small quantities disrupt human endocrine systems.

J. Lawrence Wilson, chairman of the CMA board's Research Committee, said that the project's openness — it will involve non-industry scientists and will publicly disseminate all results — “reaffirms our industry's commitment of assuring the safety of our products, even if it hurts”.

Demoted PhD accused of trying to kill whistleblower

[MUNICH] A German veterinary scientist who was last year stripped of his PhD after an investigation concluded he had manipulated some of his results, has been charged with the attempted murder of the whistleblower in the case, Guangming Xiong.

The scientist had worked at the University of Giessen on the cloning and expression of a rabbit gene encoding a binding protein for a *Pseudomonas* cytotoxin. According to the news magazine *Der Spiegel*, a series of incidents in the university's Institute of Pharmacology and Toxicology accompanied growing concern about the validity of his work during 1997.

Xiong was taken to hospital and treated in the intensive care unit in June 1997, after drinking tea laced with digitoxin in his laboratory. The accused scientist, who is currently working as a veterinarian in a private practice near Giessen, is appealing

against the derecognition of his PhD thesis, and denies the charge of attempted murder.

Scientists win highest Australia Day honours

[SYDNEY] Scientists featured more prominently than ever before in this year's Australia Day honours, winning two of five awards in the country's top category for community service: Companion of the Order of Australia.

Suzanne Cory, director since 1996 of the Walter and Eliza Hall Institute for Medical Research in Melbourne, was recognized for research into the molecular basis of cancer and the genetic “accidents” that cause cancer.

The other winner is Malcolm McIntosh, who has been dependent on dialysis since having both cancerous kidneys removed. McIntosh became chief executive of the Commonwealth Scientific and Industrial Research Organisation three years ago.

correction

A recent article quoted palaeo-anthropologist Meave Leakey as arguing that the need for tighter restrictions on the export of human fossils had been highlighted by a request from the University of Chicago (see *Nature* **396**, 504; 1998). The quotation should have referred to the Field Museum in Chicago.

Risks inherent in fetal gene therapy

Sir — We wish to express our opposition to W. French Anderson's proposed gene therapy experiments on fetuses, which are likely to result in genetic changes in fetal germline cells (*Nature* **395**, 420; 1998).

We are not opposed to fetal gene therapy in principle. However, the UK Gene Therapy Advisory Committee should be congratulated for ruling out direct injection of viral vectors into fetuses on safety and ethical grounds. We believe that Anderson's proposals constitute the first step towards intentional human germline genetic engineering, which will mainly be used for purposes of 'enhancement'. It is vital that the widely accepted prohibition on human germline genetic engineering be maintained.

The proposed experiments, intended to treat severe combined immune deficiency (SCID), are scientifically premature given current lack of success with somatic gene therapy. Furthermore, there are alternative therapies for SCID, such as neonatal bone marrow transplantation, especially where a newborn child is known to be at risk.

Although heritable changes will be, formally speaking, an unintentional consequence of the experiments, they are predicted and likely, and could therefore be considered as deliberate. Anderson is on record as supporting germline engineering for certain purposes, and has said that he intends to "force the debate" on human germline genetic engineering. These facts suggest that the proposed experiments should be treated as intentional human germline genetic engineering.

However, there is a danger that the experiments will be approved within US regulatory guidelines, because intentional germline changes are not the stated purpose of the experiment.

The prohibition on human germline genetic engineering, established in law in many countries, is well founded on social and ethical grounds. There are very few examples where existing alternatives, such as gamete donation and preimplantation

genetic diagnosis, would not provide the opportunity for families affected by genetic disease to have genetically related children. There are also the traditional options of adoption and non-parenthood. We believe that, once its use became established for medical purposes, it would be impossible to prevent the use of germline engineering for purposes of 'enhancement'.

We urge the US Recombinant DNA Advisory Committee and regulators in other countries to follow the lead of the United Kingdom, and reject Anderson's proposals.

David King*, **Tom Shakespeare†**, **Richard Nicholson‡**, **Angus Clarke§**, **Sheila McLean¶**

*GenEthics News, PO Box 6313,
London N16 0DY, UK

†Department of Sociology, University of Leeds,
Woodhouse Lane, Leeds LS2 9JT, UK

‡Bulletin of Medical Ethics, 8 Rose Cottage,
22–24 Highbury Grove, London N5 2EA, UK

§Institute of Medical Genetics, University Hospital
of Wales, Heath Park, Cardiff CF4 4FW, UK

¶School of Law, University of Glasgow,
Glasgow G12 8QQ, UK

Anderson replies — I welcome the chance to respond to the thoughtful letter of King *et al.* On one issue we absolutely agree: the use of germline genetic engineering for purposes of 'enhancement' would be a calamity for society. But concern about potential future misuse should not be allowed to prevent the legitimate development of a technology that can save lives and relieve suffering.

King *et al.* and their colleagues at the US Council for Responsible Genetics (CRG) believe that the danger is so great that no attempt at fetal gene therapy should be allowed that could result in any germline gene transfer. I disagree. A new technology that can save lives should not be rejected because of a theoretical risk. As a paediatrician I have spent my life attempting to reduce the suffering and death of children. Fetal gene therapy offers a powerful technology that could correct genetic diseases which now produce

irreversible damage before birth, so avoiding the need for alternative techniques such as gamete donation, embryo selection, abortion, adoption or non-parenthood.

King *et al.* are incorrect in stating that fetal gene therapy is "likely to result in genetic changes in fetal germline cells". Our proposal is to treat the second-trimester fetus (not the embryo), by which time all organ systems have formed. There is no evidence that gene transfer could take place into the primary germ cells even if it were desired. All the published data indicate that germline gene transfer is highly unlikely, or may not even be possible, using present techniques (see, for example, refs 1 and 2). So "heritable changes" are not "deliberate", and it is incorrect to state that "the proposed experiments should be treated as intentional human germline genetic engineering".

It is misleading for King *et al.* to state that other countries should "follow the lead of the United Kingdom, and reject Anderson's proposals". Our proposals have never been submitted to any UK body. We did not ask the US Recombinant DNA Advisory Committee for approval: we requested that a discussion of these issues should begin. The committee commended us for our actions.

The real issue is how to reap the benefits of fetal gene therapy while avoiding misuse. I believe the best defence for society is an educated public. Therefore I applaud King *et al.* and the CRG for their commendable efforts to raise public awareness of these issues. We have presented our proposals several years before we will be ready to carry out fetal gene therapy, to allow the public to evaluate our efforts at every stage. If we cannot reduce the risk of germline gene transfer to a minimal level, we will not proceed with our proposed studies.

W. French Anderson

Gene Therapy Laboratories, University of Southern
California School of Medicine, Los Angeles,
California 90033, USA

1. Ye, X. *et al.* *Hum. Gene Ther.* **9**, 2135–2142 (1998).

2. Porada, C. D. *et al.* *Hum. Gene Ther.* **9**, 1571–1585 (1998).

Visible viruses and geodesic domes

Sir — Martin Kemp's¹ article "Visible viruses" incorrectly credits Brenner and Horne with introducing negative staining and Wildy and Watson with recognizing the parallels between virus structure and Buckminster Fuller's geodesic domes. His discussion of capsid design omits key points.

H. E. Huxley's² high-resolution study of

tobacco mosaic virus made the 'outline' of the virus visible, with a clear hole down the middle, saying: "This 'outlining' technique ... is so simple and gives excellent contrast and resolution".

Caspar and Klug's³ comparison of the geodesic domes of Buckminster Fuller to the viral capsid design is discussed by Marks⁴: "Dr A. Klug and Dr J. T. Finch [*sic*] ... wrote to Fuller enclosing published reports of their discovery of the icosahedron-geodesic structuring of the polio virus. In conversations with Fuller ... they intimated that all spherical

viruses comprise geodesic arrangements of proteins in systems similar to Fuller's frequency modulated geodesic structures."

Finally, Kemp asks whether capsid design is constrained by the physical principles underlying macromolecular assemblies, by the principles governing the encoding of design in the genome, or both. The question revolves around function and inheritability — how does the virus genome protect itself in transit between cells?

A simple solution was proposed by Crick and Watson⁵, who considered how to fit the

information for a large container into a small genome. They argued that the repeated use of a single protein to form a symmetric viral capsid could require as little as one gene, making efficient use of genetic information. Today, asymmetrical designs, like those of the cytoskeleton, capture our imagination. Their routes to self-assembly, unlike the pathway described by Lewis Carroll, do not keep "one principle object in view — to preserve its symmetrical shape".

David J. DeRosier

(Director)

W. M. Keck Institute of Cellular Visualization,
Brandeis University, Waltham,
Massachusetts 02454, USA

1. Kemp, M. *Nature* **396**, 123 (1998).
2. Huxley, H. E. *Proceedings of the Stockholm Conference on Electron Microscopy* 260–261 (Stockholm, 1956).
3. Caspar, D & Klug, A. *Cold Spring Harb. Symp. Quant. Biol.* **XXVII**, 1–24 (1962).
4. Marks, R. W. *The Dymaxion World of Buckminster Fuller* (Reinhold, New York, 1960).
5. Crick, F. H. C. & Watson, J. *Nature* **177**, 473–475 (1956).

Spanish researchers defeated by the system

Sir — It is with the utmost satisfaction that I read the articles on Spanish universities *Nature* **396**, 709, 712; 1998). So far, the press, with a few exceptions which now include *Nature*, have simply echoed the encouraging words of the present and previous Spanish governments. Though many alleged 'reforms' of the last 15 years may look quite sensible, insiders can see that successive governments have promised much and delivered nothing. They claim to devote more funds to research, hoping the figures won't look too embarrassing when Spain's scientific budget is compared with those of other EU members every year.

It is true that, as the science budget has increased, so has the number of publications. It is also true that several research groups are among the best in the world in their fields. What the Government fails to mention is how many publications are by Spanish scientists working abroad, and how many are the result of international collaborations. The top groups' achievements are more due to the devotion of their members than to any government policy. The Government's zealotry in allocating more research funding has never been followed by any interest in ensuring that these funds (often from the EU rather than the Spanish budget) were properly used. There are no inspections, no yield assessments, no questions asked.

Against this background, vast sums of money have been invested in the creation of dozens of ridiculously small and unnecessary universities over the past decade. Hundreds of positions have been

created and filled — most, if not all, by local applicants with hardly enough experience to defend a master's thesis. These people are now being promoted to permanent professorships, while much more qualified candidates are humiliated and rejected on the most extraordinary grounds. In return for making these appointments, the university rectors and heads of departments ensure their own right to do exactly as they please, if anything at all.

The case of Dr Ferriz, described in your articles, would not surprise any Spanish scientist. In fact, what would be surprising is to learn that at least one Spanish university has not yet been sued. I took the University of Seville to court in 1994, for very similar reasons to those of Dr Ferriz. The outcome of the litigation, which I expect any time within the present geological era, is as unpredictable as next week's lottery results, given that the Spanish legal system is as anachronistic and bureaucratic as the one that rules the universities, and judges only rarely consider information from independent researchers.

It results in a system which is essentially unfair and leaves everybody defenceless. I personally will not attempt to return to Spain until the Government, for the first time in history, seriously decides to implement a rational science policy.

Francisco Rodríguez-Quinones

School of Biological Sciences,
University of Birmingham, Edgbaston B15 2TT, UK

Can teaching ethics make people ethical?

Sir — In a leading article you claim that "extreme misconduct tends to occur most frequently" in biomedical research (*Nature* **395**, 727; 1998). You fail to cite data to substantiate the claim that extreme misconduct in biomedical sciences is more frequent than in other sciences, but it is obvious that it is vastly more frequent than is acceptable. It is also embarrassingly obvious that more ethics is probably taught in medical faculties than in any other scientific discipline. It is hard to find a medical school in the developed world without at least some formal medical ethics education. But I doubt if any mathematics or physics departments offer courses in 'mathematical ethics' or 'physical ethics'.

Those of us who teach medical ethics must therefore ask ourselves the hard question whether we are really teaching what we ought to teach, and whether 'ethical people' can really be produced by academic instruction. Classes on ethics of genetic engineering, cloning, brain death and when the embryo becomes a person

can be exciting and intellectually stimulating but they don't necessarily make physicians, nurses and biomedical researchers more honest.

Frank J. Leavitt

Medical Ethics Centre, Faculty of Health Sciences,
Ben Gurion University of the Negev,
Beer Sheva, Israel
email: yeruham@bgumail.bgu.ac.il

Glaciation: the snowball theory still holds water

Sir — D. M. Williams *et al.*¹ propose a mechanism² whereby an initial high obliquity for the Earth could have rapidly changed to its present low value of ~23.5° near the end of the Proterozoic eon. Following G. Williams³, they note that the mean annual insolation would have been lower at the Equator than at the poles if the Proterozoic obliquity exceeded 54°. They infer that low-latitude glaciation observed near the beginning and end of the Proterozoic could be explained without the need for an ice-covered 'snowball' Earth^{4,5}.

However, neither Williams *et al.*¹ nor the accompanying News and Views article⁶ discuss the implications of high obliquity for glaciation *per se*. The basic requirement for glaciation is a net accumulation of winter snow after summer melting. High obliquity enhances seasonality^{3,7}, creating very cold winters with reduced snowfall and very hot summers with maximal melting.

The recognition that glaciation is favoured by cool summers, not cold winters, is the crucial difference between the Milankovitch theory and the much earlier Croll theory of orbital forcing⁸. High obliquity has the greatest effect on seasonality at the poles, but insolation during the warm seasons at the equator is equally high irrespective of obliquity.

Whatever merit it has for the early Earth¹, high obliquity seems insufficient as an explanation for low-latitude glaciation⁷ and glaciating the poles as a means of reducing a high obliquity¹ is highly improbable.

Paul F. Hoffman, Adam C. Maloof

Department of Earth & Planetary Sciences,
Harvard University, Cambridge,
Massachusetts 02138, USA

1. Williams, D. M., Kasting, J. F. & Frakes, L. A. *Nature* **396**, 453–455 (1998).
2. Bills, B. G. *Geophys. Res. Lett.* **21**, 177–180 (1994).
3. Williams, G. E. *Earth-Sci. Rev.* **34**, 1–45 (1993).
4. Kirschvink, J. L. in *The Proterozoic Biosphere* (eds Schopf, J. W. & Klein, C.) 51–52 (Cambridge Univ. Press, New York, 1992).
5. Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. *Science* **281**, 1342–1346 (1998).
6. Bills, B. G. *Nature* **396**, 405–406 (1998).
7. Hunt, B. G. *J. Meteorol. Soc. Jpn.* **60**, 309–318 (1982).
8. Imbrie, J. & Imbrie, K. P. *Ice Ages* (Harvard Univ. Press, Cambridge, 1979).

From *Pan* to pandemic

Robin A. Weiss and Richard W. Wrangham

Further evidence that HIV-1 originally came from chimpanzees bears on a variety of issues – the evolution of AIDS viruses, disease transmission from animals to humans, and chimpanzee conservation and welfare.

The origin of human immunodeficiency virus type 1 (HIV-1), the retrovirus that is the main cause of AIDS, has been a puzzle ever since it was discovered by Barré-Sinoussi and her colleagues¹ in 1983. On page 436 of this issue², Gao *et al.* provide the most persuasive evidence yet that HIV-1 came to humans from the chimpanzee, *Pan troglodytes*, which harbours a related simian immunodeficiency virus, SIVcpz. Moreover, by genetic typing of mitochondrial DNA, it is clear that the three strains of SIVcpz most closely related to HIV-1 come from a single subspecies, *Pan troglodytes troglodytes*, which lives in the same part of central Africa where AIDS is thought to have arisen.

There are three principal groups of HIV-1: the main (M) group comprises the majority of subtypes that have spread across the world; an outlier (O) group is found in Cameroon, Gabon and Equatorial Guinea; and a new group (N) was last year identified in two people in Cameroon³. HIV genetic sequences from a Congolese serum sample, originally taken in 1959, showed that the diversification of all the M subtypes has occurred in very recent times, say the past 50 years⁴. Gao and colleagues' evolutionary analysis of the M, N and O groups in comparison to SIVcpz isolates indicates that these HIV-1 groups represent three separate transfers from chimpanzees to humans (such inter-species transmissions of infectious disease are known as zoonoses).

It has been shown^{5,6} previously that several strains of HIV-2 in West Africa were independently derived from SIVsm, the virus endemic in the sooty mangabey monkeys of that region. The other human retroviral pathogen, human T-lymphotropic virus type 1 (HTLV-1), has also originated more than once from related simian viruses, including STLV-1 of chimpanzees^{7,8}. And a new paper⁹ reporting a parallel zoonosis to that of Gao *et al.*² has demonstrated that human pygmies in Gabon are infected with a strain of HTLV-1 that is virtually indistinguishable from STLV-1 of mandrills living in the same forest. Finally, a survey of people who handle primates in captivity recorded several instances of foamy retrovirus transmission and one transfer of SIVmac¹⁰, the strain that causes AIDS in macaques.



Figure 1 Chimpanzee orphan, chained in a junk car lot in Cameroon. The rest of its family had been shot and butchered for the bushmeat trade.

So cross-species transmission of primate retroviruses to humans through hunting or husbandry probably occurs relatively frequently, although — thankfully — the subsequent epidemic spread in the human population is a much rarer event. Such zoonoses have heightened the concern that retroviruses might also cross to humans from more distant species, such as pigs, if their tissues were used for xenotransplantation¹¹.

Gao *et al.*² also show that a chimpanzee called Noah, from which a highly divergent strain, SIVcpzANT, was isolated, actually belongs to the different subspecies *Pan troglodytes schweinfurthii* (the names and distributions of all four subspecies are shown in the map on page 438). Gao *et al.* argue that SIVcpz is an ancient chimpanzee infection that has co-evolved with its host during speciation. They previously followed the same logic for the diversification of SIVagm among the various subspecies of African green monkey¹². But these monkeys, like the chimpanzee subspecies, have largely non-overlapping

habitats, so it is also conceivable that SIVs were more recently introduced and owe their diversity to geography rather than co-evolution.

Indeed, the ability of SIV to jump host species, together with the demonstration by Gao *et al.* that one strain of SIVcpz is (like HIV-1E in Thailand) derived from genetic recombination between two distinct parent strains, suggests that the co-evolution hypothesis needs to be viewed with a little caution. To bolster Gao and colleagues' conclusion that chimpanzees are a natural reservoir population for the precursor of HIV-1, serological surveys for the prevalence of SIVcpz infection are required in wild populations in Africa.

The ramifications of the new findings² extend well beyond the evolution of HIV-1, and zoonotic transmission of disease, to chimpanzee conservation and welfare. Like the other great apes, chimpanzees are in trouble. In most of the 21 African countries where they live in the wild, their numbers are plummeting — partly because of habitat loss, and partly because ape-hunting has become big business (Fig. 1). Hunters are paid by timber companies to provision logging camps, while the kills of freelance hunters are traded as far as the cities. Ape meat can fetch premium prices in middle-class restaurants.

This 'ape bushmeat crisis' cannot last long. Numbers are hard to come by, but some estimates of the annual chimpanzee kill are in the thousands, which is an unsustainably high figure when as long as a decade ago the world population was estimated at around 200,000. In the face of such an onslaught, populations of the large, slowly reproducing apes will be swiftly eradicated¹³.

Hunters dismember chimpanzees with primitive butchery, and so expose themselves to the risk of zoonotically transmitted disease. Conservationists, therefore, are debating the merits of a campaign that would publicize the disease-transmitting dangers of treating apes as food. Some are concerned that this could backfire, and result in intensified slaughter of the primates, whereas others think the situation could hardly get worse. Later this month, a meeting hosted by the American Zoological Association aims to turn discussion into action. In the long term, apes will survive in the wild only where they are not eaten, a prospect that depends on the acceptance of new values by which apes are treated as too human-like to kill or eat. Many believe that new ethics are also needed in the scientific world. To some AIDS researchers, the use of chimpanzees is constrained merely by their high cost and scarcity¹⁴. By 1996, 198 chimpanzees had been experimentally infected with HIV in six of the major research institutions in the United States¹⁵, although HIV seldom causes disease in these animals.

Worldwide, there are now about 35 million carriers of HIV-1, and with the increased

KARL AMMANN

evidence that the virus is linked to SIVcpz there may be calls for more use of chimpanzees in AIDS research. But according to a growing constituency, evidence of ape possession of human-like cognition and emotions should reduce our willingness to infect them, and condemn them to solitary confinement for life (which lasts for 60 years or more)¹⁶.

The approach of Gao *et al.* may prompt fresh thinking. Fieldwork with free-living apes provided information for the evolutionary history of chimpanzee subspecies that was critical to their results. Further fieldwork linking demographic data to biomedical monitoring could be even more productive. The four chimpanzee subspecies have deep evolutionary roots, and may yield further types of SIV beyond the two already identified. These are best studied in the place where the host-virus systems evolved. Chimpanzees in captivity are mostly taken from the wild before they become sexually active, and so rarely harbour SIV. Infection can be expected at higher frequencies in wild adults¹⁷, and such individuals would offer research opportunities that simply do not pertain in captivity, where experimenters cannot know how virus strains match to genetic lineages.

Two other ways forward are these. Field researchers intermittently find fresh corpses of individual apes, often with well-documented life-histories, which offer untapped possibilities for virological analysis. And DNA studies of faeces might allow non-invasive monitoring of virus infection.

Biomedical researchers have the prospect of collaborating with fieldworkers in a synthesis that would benefit conservation. The link between HIV-1 and SIVcpz may open a door for research that helps both humans and apes. □

Robin A. Weiss is in the Windeyer Institute, University College London, 46 Cleveland Street, London W1P 6DB, UK.

e-mail: robinw@icr.ac.uk

Richard W. Wrangham is in the Department of Anthropology, Harvard University, Cambridge, Massachusetts 02138, USA.

e-mail: wrangham@fas.harvard.edu

1. Barré-Sinoussi, F. *et al.* *Science* **220**, 868–871 (1983).
2. Gao, F. *et al.* *Nature* **397**, 436–441 (1999).
3. Simon, F. *et al.* *Nature Med.* **4**, 1032–1037 (1998).
4. Zhu, T. *et al.* *Nature* **391**, 594–597 (1998).
5. Gao, F. *et al.* *Nature* **358**, 495–499 (1992).
6. Chen, Z. *et al.* *J. Virol.* **71**, 3953–3960 (1997).
7. Koralknik, I. J. *et al.* *J. Virol.* **68**, 2693–2707 (1994).
8. Voevodin, A. *et al.* *Virology* **238**, 212–220 (1997).
9. Mahieux, R. *et al.* *J. Virol.* **72**, 10316–10322 (1998).
10. Heneine, W. *et al.* *Nature Med.* **4**, 403–407 (1998).
11. Weiss, R. A. *Nature* **391**, 327–328 (1998).
12. Sharp, P. M., Robertson, D. L. & Hahn, B. H. *Phil. Trans. R. Soc. B* **349**, 41–47 (1995).
13. Wilkie, D. S. & Carpenter, J. *Biodivers. Conserv.* (in the press).
14. Letvin, N. L. *Science* **280**, 1875–1880 (1997).
15. National Research Council *Chimpanzees in Research: Strategies for their Ethical Care, Management, and Use* (NRC, Washington, DC, 1996).
16. Singer, P. & Cavalieri, P. (eds) *The Great Ape Project* (St Martin's, New York, 1994).
17. Jolly, C. J. *et al.* *J. Med. Primatol.* **25**, 78–83 (1996).

Condensed-matter physics

Return of the itinerant electron

David Ceperley

On page 412 of this issue, Young *et al.*¹ demonstrate magnetic polarization in a dilute, three-dimensional electron gas at a temperature of 600 K. Years of theoretical debate connect this intriguing result to a long-sought mechanism for ferromagnetism in metals first conjectured by Bloch², who suspected that the 'electron gas' is susceptible to magnetic ordering at low density.

What is an electron gas and why is it interesting? The idealized homogeneous gas (also known as jellium or the one-component plasma) has a long history going back to the discovery of the electron. It was used by the early quantum pioneers to model the valence electrons in a solid; they replaced the ions by a rigid uniform background of positive charge. This is arguably the most important model in condensed-matter physics because it is routinely used as a reference state in most realistic calculations of electronic structure. The valence electrons are the outer electrons of the atom, which determine how it interacts with its environment, so understanding the properties of the electron gas is significant. In most metals, the density is high enough that the electrons move almost independently and form a 'normal Fermi liquid' of mobile electrons (Fig. 1) with equal numbers of electrons with up and down spin. But at low den-

sity the potential energy dominates the kinetic energy, and the electron 'gas' freezes into the 'Wigner crystal'³, with the electrons localized on lattice sites. This dependence on density is the exact opposite found in systems with cores, such as atoms, where it is high not low density that favours a lattice structure.

In 1929, Bloch noticed that within the 'Hartree-Fock' approximation (that is, assuming the electrons are in independent states), the spins of the electrons spontaneously align, and he suggested this as an explanation of ferromagnetism. This spin polarization should happen at a density just lower than that found in some alkali metals, such as potassium, rubidium or caesium. But at these densities one cannot ignore the fact that electrons move in a correlated way. In the Stoner model⁴ introduced in the 1930s, the electron-electron Coulomb interaction is replaced by a constant repulsive energy between opposite spin electrons, to account for this correlation. The spins only partially polarize and they do so at a lower density. Many theorists have concluded from this and other calculations that the electron gas would never be ferromagnetic⁵. Calculations at an intermediate density are difficult because very small energy differences are important and any approximation has to treat the various 'phases' of the gas with equal accuracy.

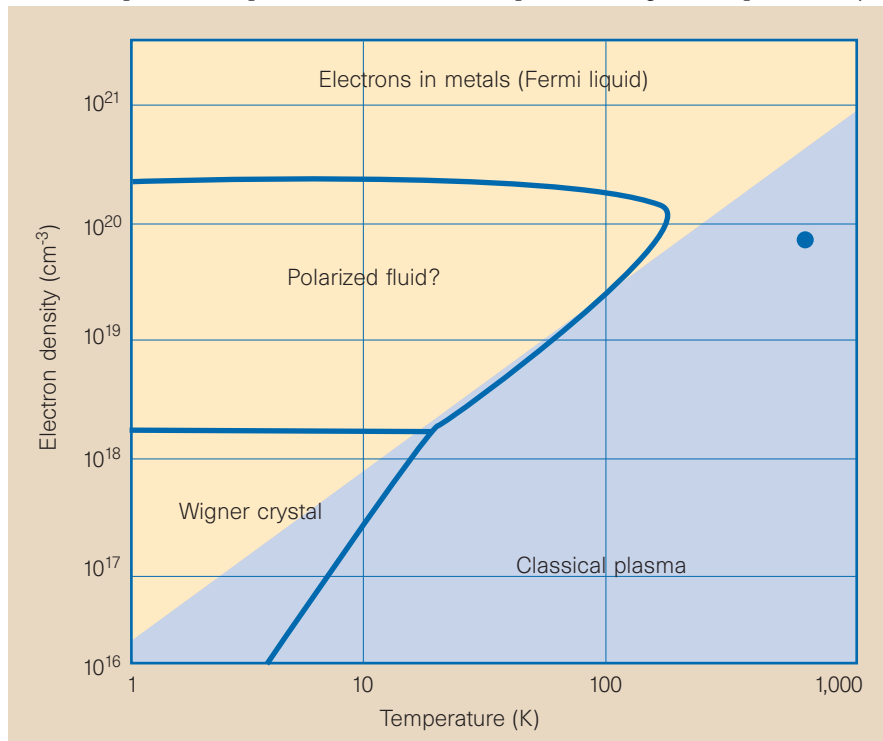


Figure 1 Phase diagram of the electron gas. The two colours divide the classical (blue) from the quantum (yellow) regimes. The phase transition boundaries are estimates from ref. 6. The dot is the transition temperature measured by Young *et al.*¹.

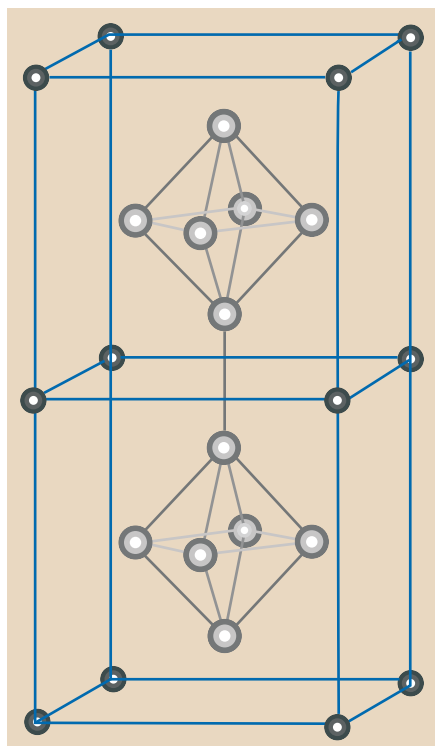


Figure 2 Crystal structure of CaB_6 . The calcium atoms are in black and the boron atoms in grey.

The most reliable calculations⁶ have used a computer simulation method known as quantum Monte Carlo. These calculations predict that electrons will spin-align at a density less than $2 \times 10^{20} \text{ cm}^{-3}$ and crystallize at a density of $2 \times 10^{18} \text{ cm}^{-3}$ into the body-centred cubic lattice of the Wigner crystal. The magnetic transition is predicted to be gradual with the spin polarization building up slowly as the density drops, just as in the Stoner model. Such computer calculations have only been done at zero temperature; so the estimate of the transition temperature shown in Fig. 1 is a combination of the computer results and the Stoner model. These predictions have remained untested because of the difficulty of achieving a physical realization of the electron gas at low densities.

Young *et al.*¹ report ferromagnetism in a system that is plausibly modelled by the electron gas. They take calcium hexaboride (CaB_6 , a semi-metal) and replace a few of the calcium atoms with lanthium atoms (Fig. 2). Each impurity atom adds an electron to the conduction band. By changing the impurity concentration they vary the effective electron density just as in doped silicon semiconductors. Remarkably, the authors observe spontaneous partial magnetization for a range of densities, in agreement with the predictions of the electron gas calculations. Several analogous compounds, for instance strontium hexaboride, show the same ferromagnetism. However, the transition temperatures at which ordering occurs are a factor of three higher than the computer estimates. In order

to interpret the experiment it is necessary to correct for 'band effects', because the electrons are not free but move inside a cubic lattice. One important correction is that conduction electrons move in states oriented to the cubic crystal axes, producing an extra quantum number not considered in the electron-gas model. In addition, the effective mass of conduction electrons in CaB_6 is half their vacuum value⁷ and depends on direction, and the other electrons screen the electronic charge. Such effects enhance the natural tendency of the electron gas to spin align. These corrections are so important that detailed calculations will be needed to establish the role of itinerant electrons in the observed ferromagnetism.

Such spin-spin correlations are ubiquitous in strongly correlated Fermi liquids and can dominate how these systems react to

their surroundings. The clearest example is superfluid helium-3, which although not ferromagnetic, is nearly so. A more complicated and disputed example is of spin correlations in the high-temperature superconductors. Experiments with hexaborides could help the essence of the ferromagnetic electron to be revealed. □

David Ceperley is in the Department of Physics and NCSA at the University of Illinois, Urbana-Champaign, Illinois 61821, USA.

e-mail: ceperley@ncsa.uiuc.edu

1. Young, D. P. *et al.* *Nature* **397**, 412–414 (1999).
2. Bloch, F. *Z. Phys.* **57**, 545–555 (1929).
3. Wigner, E. P. *Phys. Rev.* **46**, 1002–1011 (1934).
4. Stoner, E. C. *Proc. R. Soc. Lond. A* **165**, 372–414 (1938).
5. Herring, C. in *Magnetism* Vol. IV (eds Rado, G. T. & Suhl, H.) 9–31; 71–80 (Academic, San Diego, 1966).
6. Ceperley, D. M. & Alder, B. J. *Phys. Rev. Lett.* **45**, 566–569 (1980); *Int. J. Quant. Chem.* **16**, 49–61 (1982).
7. Massida, S. *et al.* *Z. Phys. B* **102**, 83–89 (1997).

Apoptosis

A cellular poison cupboard

William C. Earnshaw

Apoptotic cell death is spectacular — one moment a cell looks peaceful and happy, the next it enters a violent programme of cytoplasmic blebbing worthy of the death throes of an actor in a B-movie. One fascinating aspect of apoptosis is that virtually all cells contain latent forms of the molecules that trigger this destructive burst. A report by Susin *et al.*¹ on page 441 of this issue supports the idea that several of these key death factors normally reside in the space between the inner and outer membranes of the mitochondrion.

Apoptosis is driven by two classes of highly specialized proteases known as caspases (cysteine aspartases). When properly assembled on a molecular scaffold, 'initiator' caspases can be activated by self-cleavage. 'Effector' caspases are then activated downstream of the initiator caspases in an amplifying cascade. Mitochondria come into play because they intervene between these two classes of caspase in many apoptotic pathways. For example, active caspase-8 (an initiator caspase) can act either directly on downstream effector caspases or indirectly on mitochondria, in both cases releasing a cocktail of lethal factors that amplify the death response.

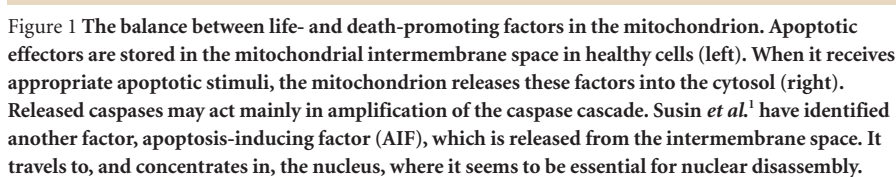
The first such mitochondrial factor to be described binds to a cytoplasmic scaffolding protein called Apaf-1. Binding of the mitochondrial factor allows Apaf-1 to form a ternary complex with, and activate, the initiator procaspase-9 (ref. 2). Active caspase-9 then turns on downstream effector caspases, unleashing the death cascade. The announcement that this mitochondrial factor was none other than cytochrome *c*, a component of the respiratory chain that is

normally found in the mitochondrial intermembrane space³, was greeted with incredulity — it was a great surprise that such a well-understood molecule could lead such a Jekyll-and-Hyde existence.

Susin and colleagues¹ now extend previous experiments to show that purified mitochondria can release another factor that induces apoptotic morphological changes in nuclei *in vitro*⁴. They have termed this factor apoptosis-inducing factor (AIF), and find that it is a widely expressed flavoprotein. Found in the mitochondrial intermembrane space, AIF has sequence homology to bacterial oxidoreductases (electron acceptors and donors). The AIF is released from mitochondria by apoptosis-inducing stimuli, allowing it to travel to, and accumulate in, the nucleus.

Susin *et al.* provide compelling evidence that AIF deserves a place on the centre stage, along with other well-known components of the apoptotic machinery. During apoptotic execution, caspases act both as executioners that cleave key proteins, and as executives that turn cell-survival pathways off and death-promoting activities on. One such death-promoting factor is a nuclease called caspase-activated DNase⁵ (CAD; also known as caspase-activated nuclease and DNA fragmentation factor 40 kD). Purified, recombinant CAD can, on its own, induce the collapse and hypercondensation of the chromatin against the nuclear periphery — events diagnostic of apoptotic execution. Yet CAD is not essential for these final events of nuclear disassembly⁶, indicating that other factors are involved.

Sure enough, when Susin *et al.* added recombinant AIF to isolated nuclei, they saw chromatin condensation at the nuclear



Both of these studies from Susin and col-

389

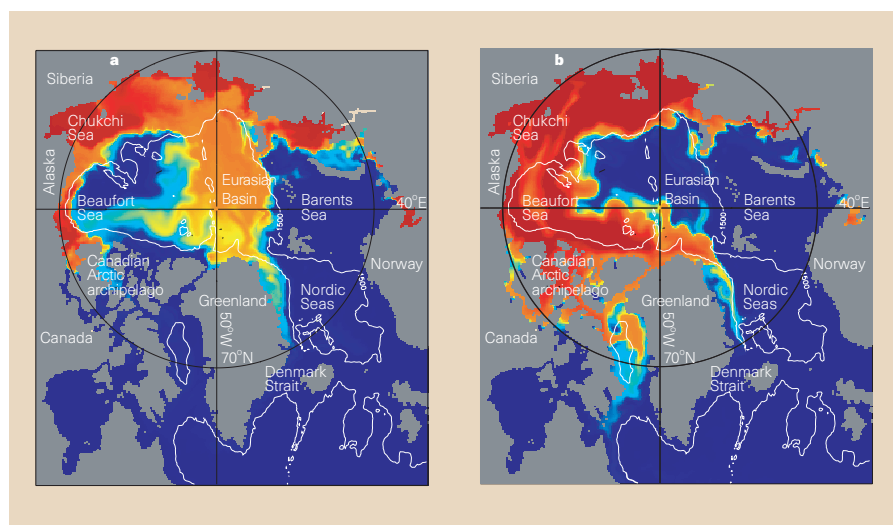


Figure 1 Simulating the effect of climate on freshwater distributions in the Arctic. Both a and b show the freshwater tracer distribution (light tones) at 0–20 m depth for 20-year integrations of the Advanced Arctic Ocean Model of the Naval Postgraduate School, Monterey, using repeated (a) 1979 and (b) 1990–94 atmospheric forcing from the European Centre for Medium-range Weather Forecasts (W. Maslowski, NPS, personal communication, and ref. 14). Each of eight freshwater sources has been followed as a separate ‘active’ tracer throughout the Arctic Ocean. The result supports field observations in suggesting a retraction and eastward redistribution of the surface freshwater layer from the Eurasian Basin during the anomalous atmospheric regime of the 1990s.

enough to be traceable through a gappy observing system and stand out against a still-shaky climatology. So what has been happening in the Arctic Ocean? The abstracts⁷ of the latest session amplify the results of a workshop on the same topic held in 1997 (ref. 8).

By the late 1980s and early 1990s, an extreme amplification of the NAO had carried a warmer, fresher and probably stronger transport of Norwegian Atlantic Water north to the Fram Strait and Barents Sea⁴. Entering the Arctic, the Atlantic-derived sublayer shoaled and warmed by up to 2 °C in the Eurasian Basin and extended in distribution by about 20% (ref. 8); accordingly, the first of the SCICEX surveys showed that the mutual front between waters of Pacific and Atlantic origin had shifted from the Lomonosov to the Alpha-Mendelev Ridge. This strengthened invasion of Atlantic water was successfully modelled.

At lesser depths, the cold halocline (freshwater cap), which had acted to insulate the sea-ice from the warm Atlantic layer below, dwindled away in the Eurasian Basin⁹ with profound effects on the surface energy- and mass-balance of sea-ice in that region. This change seems to have stemmed from the eastward diversion of Russian rivers in response to the altered atmospheric circulation, and was apparently captured by the most powerful of the coupled Arctic models (Fig. 1).

Arctic-wide, the pattern of atmospheric pressure and ice drift shifted anticlockwise in a similar sense. The altered pressure gradients drove an increased flux of ice south to Nordic Seas⁴, and the southward recirculation of warmth from the eastern Fram Strait even affected the Denmark Strait overflow

far to the south¹⁰. Most recently, a warmer, fresher surface layer, thinning sea-ice and a record extent of open water have affected the western Arctic¹¹.

Yet these discoveries represent only a ‘stage one description’ of events, reminiscent of our understanding of variability in the equatorial Pacific before the Tropical Ocean Global Atmosphere (TOGA) programme. We have observed unprecedented and system-wide change through the ocean–atmosphere system of the Arctic. We believe we can distinguish elements that seem locally generated from others imposed from outside. We have correlated these changes with local and external climatic forcing. And we can document a potential for global as well as regional effects on climate — acting locally through the effect of a changing ocean stratification and ice cover on surface heat flux, and remotely through the variable export of fresh waters and heat to the headwaters of the global thermohaline circulation.

Following the TOGA example, setting these observations into a clean physical context will involve teasing out the processes and mechanisms at work, particularly those which determine how the inner workings of the Arctic system might control or modify the imposed climatic signal. This necessary preoccupation with process is already well under way. The Surface Heat Budget of the Arctic Ocean project of the Arctic System Science (ARCSS) programme completed its one-year field phase in the Beaufort Sea in October 1998. The Canadian–US Joint Ocean Ice Study has for the first time addressed the processes of exchange through

the Canadian Arctic archipelago. A second ARCSS study on shelf–basin interactions of the Chukchi–Beaufort margin is just getting underway. New tracer techniques can respectively partition the freshwater balance into its river, ice-melt and Pacific components, and reveal the pathways and time-scales of water-mass invasions, including recent evidence of Pacific waters spreading to the open North Atlantic.

Constraints on further progress remain, however, and three in particular recurred at the meeting. First, by the year 2000 the submarine fleet will be much reduced, the 637 *Sturgeon* class will be decommissioned, and the Arctic capability of the replacement 688 *Los Angeles* class boats will not yet be proven. The cruise by the *Hawkbill* in April 1999 is likely to be the last to carry scientists for some time. Second, it is already apparent in the Arctic, where barotropic (depth-independent) flows dominate, that small-scale topography may drive narrow but important filaments of boundary flow, providing testing conditions for both observing systems and models. Finally, the sub-Arctic seas can clearly act either as the source of Arctic change or as a conduit through which such change may have global effects. In other words, the sub-Arctic is part of the system under study and the observing scheme must cover both.

That said, oceanographers have already come a long way in capturing the startling changes in the Arctic Ocean. If the longest proxies for the dominant forcing prove valid, these will have included effects of the most positive winter NAO index for over 500 years¹². The physical and chemical evidence seems mutually consistent, and it looks as if the key features can be modelled. The emphasis on the freshwater balance seems particularly apt given the global importance that coupled models currently assign to relatively minor changes in freshwater distribution at high latitudes¹³.

The new process studies have their long-term context, and agencies such as the US Global Change Research Program and the Fifth Framework Programme of the European Commission seem to appreciate that events in the Arctic have global implications. We have come a long way in the Arctic, and should go much further. □

Bob Dickson is in the Environment Group, Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Lowestoft, Suffolk NR33 0HT, UK.
e-mail: r.r.dickson@cefas.co.uk

1. Manabe, S. & Stouffer, R. J. *Nature* **364**, 215–218 (1993).
2. Environmental Working Group Joint US–Russian Atlas of the Arctic Ocean, *Oceanography Atlas for the Winter Period CD-ROM* (Natl Ocean Data Center, Washington, DC, 1997).
3. Thompson, W. J. & Wallace, J. W. *Geophys. Res. Lett.* **25**, 1297–1300 (1998).
4. Dickson, R. R. *et al.* *J. Clim.* (submitted).
5. Walsh, J. E., Chapman, W. L. & Shy, T. L. *J. Clim.* **9**, 480–486 (1996).
6. Hurrell, J. W. *Geophys. Res. Lett.* **23**, 665–668 (1996).
7. Maslowski, W. & Schlosser, P. (convenors) *EOS (Suppl.)* **79**, 398–400; 413–415; 434–440 (1998).
8. Morison, J., Aagaard, K. & Steele, M. *Rep. Study Arctic Change*

- Workshop (Rep. No. 8, Arctic System Science, Seattle, 1998).
9. Steele, M. & Boyd, T. J. *Geophys. Res.* **103**, 10419–10435 (1998).
10. Dickson, B., Meincke, J., Vassie, I., Jungclauss, J. & Osterhus, S. *Nature* **397**, 243–246 (1999).
11. McPhee, M. G., Stanton, T. P., Morison, J. H. & Martinson, D. G. *Geophys. Res. Lett.* **25**, 1729–1732 (1998).
12. Stockton, C. W. & Glueck, M. E. *Long-term Variability of the North Atlantic Oscillation (NAO)* (Annu. Mtg. Am. Met. Soc., 1999).
13. Rahmstorf, S. & Ganopolski, A. *Clim. Change* (in the press).
14. Maslowski, W. *et al.* *EOS Trans.* **79**, 414 (1998).

Neurobiology

Striving for coherence

Wolf Singer

To explain how the brain interprets the world, Descartes postulated that we have a single centre — the pineal gland — where all sensory signals converge and are evaluated jointly, where decisions are reached and future actions planned. But progress in neurobiology has forced us to adopt a different view — that nowhere do we have a single centre to evaluate or coordinate computations. So how are the results of many parallel computations bound together to permit coherent perception and action? On pages 430 and 434 of this issue, Rodriguez *et al.*¹ and Miltner *et al.*² address this question by investigating the temporal coherence of electroencephalographic (EEG) signals recorded from people performing cognitive tasks.

To represent perceptual objects or motor programmes in systems that do not converge, the idea is that a particular content is represented by jointly activating an assembly of cells, rather than by the response of an individual, highly specialized neuron³. The advantage is that an almost unlimited number of different assemblies — each representing different contents — can be generated, because subsets of cells drawn from the large (but limited) pool of functionally specialized neurons in the cerebral cortex can be dynamically regrouped. However, this mechanism must allow the neurons that participate in a particular assembly to be identified unambiguously. The responses of all the neurons in a particular assembly must then be labelled, to ensure that they are processed jointly at subsequent stages.

Based on theoretical considerations⁴ and data from multi-electrode recordings in the visual cortex⁵, it has been proposed that responses are bound together and labelled by synchronized firing of the individual neurons with a millisecond precision. The assumption is that such synchronized activity summates more effectively than non-synchronized activity in the target cells at subsequent processing stages. If so, synchronization could increase the effect that a selected group of neurons has on other populations with great temporal specificity. Jointly raising the effect of a subset of responses is equivalent to functional binding because it favours further joint processing of the selected responses.

The EEG studies by Rodriguez *et al.*¹ and Miltner *et al.*² were designed to test the

predictions derived from this hypothesis. Signals from the EEG reflect, with high temporal resolution, the activity of many neurons in the cortical area beneath the electrodes. This prevents precise localization of neuronal responses, but facilitates the detection of synchronous activity. The responses of distributed neurons summate effectively — and then give rise to measurable fluctuations of the EEG signal — only if they are well synchronized. Typically, EEG fluctuations

are periodic and cover a broad frequency range (from less than 2 to greater than 60 Hz), indicating that neurons in the cerebral cortex can oscillate synchronously in various frequency bands.

The new studies^{1,2} concentrate on the so-called gamma frequency range, which is centred around 40 Hz. The authors selected this frequency because precise synchronization of neuronal responses is often associated with oscillations in the 40-Hz range⁵. But it is not trivial to detect these oscillations. Synchronized firing is usually short-lived (100–300 ms) and, except for an initial component, the oscillations are not phase-locked to the stimulus because they are self-generated. This prevents averaging and requires a method that allows brief bursts of synchronous activity to be detected in single trials. New analytical tools have made it possible to confirm that these transient gamma-oscillations exist in the human brain, and



Figure 1 'Mooney' faces used by Rodriguez *et al.*¹. Mooney faces are easily recognizable when viewed upright, but not when inverted. The authors used them to show that, in humans, neurons oscillating at a frequency of 40 Hz can synchronize across different regions of the brain. Miltner *et al.*² asked subjects to associate visual and tactile stimuli, and found similar coherence of oscillations at 40 Hz.



100 YEARS AGO

The vexed question as to the exact meaning of the phrase "one hour after sunset and one hour before sunrise" in the Local Government Act, 1888, referring to the lighting of bicycle lamps, was settled from a legal point of view in a Divisional Court on Thursday last. It had been held that sunset at Greenwich was meant, and the Bristol justices convicted a cyclist for riding a bicycle without a light an hour after sunset thus defined. The alleged offence was committed on August 19, 1898, at 8.15 p.m., which was less than an hour after sunset at Bristol, but more than an hour after sunset at Greenwich. An appeal was made against the decision of the Bristol magistrates; and at Thursday's Court the appeal was allowed, and the conviction quashed, their Lordships holding that the phrase in the Act referred to must not be understood to mean Greenwich time, but local time.

From *Nature* 2 February 1899.

50 YEARS AGO

The medical, social and economic problems created by a rapidly ageing population were the subject of a symposium ... at the British Association meeting at Brighton on Friday, September 10, 1948. The subject was introduced by Sir Ernest Rock Carling, who reminded his audience that the great majority of the elderly are healthy and independent – they outnumber the ailing and decrepit by more than 30 to 1 – and for them the most pressing problem is how to maintain to the end of their days the standard of living of their working life. The time has come for the elderly to revolt against the convention, based on sociological, not biological, grounds, that there is a fixed retirement-age beyond which they are unfit for further work. It is a truism that chronological age is no guide to capacity in the individual; but what we want to know is, how far the increased expectation of life in the last fifty years has extended working-capacity. ... Ageing begins at birth, and an athlete is 'old' at thirty-five. ... Laziness, Sir Ernest maintained, is at the bottom of much facile acceptance of 'too old at forty, fifty, or sixty'. If the Civil servant must retire, why does not the clergyman? If the barrister ceases to practice, why not the judge?

From *Nature* 5 February 1949.

to establish close relations between them and cognitive processes^{6,7}.

Rodríguez *et al.* and Miltner *et al.* now show that local oscillatory responses can synchronize across different cortical areas – the time course and topological distribution of synchronization showed a high degree of task-related specificity. These results from humans closely resemble those obtained by intracortical recording from cats⁸. Rodríguez *et al.*¹ asked people to inspect pictures that could, on occasion, be recognized as a face (Fig. 1). They found that scrutinizing the pictures was associated with increased gamma activity over cortical regions known to be involved in visual processing. But precise phase-locking of these oscillations across cortical areas occurred only when the subjects identified a face. This state of heightened synchrony was transient. It dissolved shortly before the subjects responded by pressing a key, giving way to a second episode of phase-locked gamma activity associated with the motor response, which had a different topological distribution. The authors suggest that such dynamic changes in the phase relations between spatially distributed oscillating groups of neurons could reflect the transient formation of assemblies that are bound by synchrony and represent first perception of the stimulus and then the motor programme.

Miltner *et al.*² tested the hypothesis that if people learn the association between a visual and a tactile stimulus, they should form a neuronal assembly comprising cells responsive to the visual and the tactile stimuli, respectively. The authors observed a marked increase of gamma activity after the visual stimulus was presented. Notably, they also found a selective increase of gamma coherence between the visual cortex and the cortical area representing the hand that had received the tactile stimulus. This coherence must have developed as a consequence of conditioning, because it disappeared when the learnt association was lost (that is, after a sequence of visual stimuli not connected to a tactile stimulus were presented). Moreover, such coherence was not seen for signals from

the area representing the non-conditioned hand. Finally, the increased coherence was confined to a narrow band around 40 Hz, supporting the idea that synchronization of oscillations in this frequency range is involved in cognitive processes.

The new results provide further evidence that synchronization might allow the selective association of distributed neurons. However, neither these nor previous studies have demonstrated this in the vertebrate brain, because we cannot yet disrupt synchronization in the relevant frequency band without affecting other response variables. But a functional role for synchronization has been demonstrated in the insect olfactory system⁹. There is also indirect evidence from psychophysical studies, which suggests that the brain binds responses together and interprets them as related if they are made synchronous by synchronizing the stimuli that evoke them^{10–12}. If the brain interprets responses as related when they are made synchronous by internal mechanisms (as was the case in the people studied by Rodríguez *et al.*¹ and Miltner *et al.*²), gamma oscillations could well be the mechanism that binds neurons into functionally coherent assemblies.

Wolf Singer is at the Max Planck Institute for Brain Research, Deutschordenstrasse 46, D-60528

Frankfurt/Main, Germany.

e-mail: singer@mpih-frankfurt.mpg.de

1. Rodríguez, E. *et al.* *Nature* **397**, 430–433 (1999).
2. Miltner, W. H. R., Braun, C., Arnold, M., Witte, H. & Taub, E. *Nature* **397**, 434–436 (1999).
3. Hebb, D. O. *The Organization of Behavior* (Wiley, New York, 1949).
4. von der Malsburg, C. *Ber. Bunsenges. Phys. Chem.* **89**, 703–710 (1985).
5. Gray, C. M., König, P., Engel, A. K. & Singer, W. *Nature* **338**, 334–337 (1989).
6. Tallon-Baudry, C., Bertrand, O., Delpuech, C. & Pernier, J. *J. Neurosci.* **17**, 722–734 (1997).
7. Tallon-Baudry, C., Bertrand, O., Peronnet, F. & Pernier, J. *J. Neurosci.* **18**, 4244–4254 (1998).
8. Roelfsema, P. R., Engel, A. K., König, P. & Singer, W. *Nature* **385**, 157–161 (1997).
9. Stopfer, M., Bhagavan, S., Smith, B. H. & Laurent, G. *Nature* **390**, 70–74 (1997).
10. Leonards, U., Singer, W. & Fahle, M. *Vision Res.* **36**, 2689–2697 (1996).
11. Alais, D., Blake, R. & Lee, S.-H. *Nature Neurosci.* **1**, 160–164 (1998).
12. Usher, M. & Donnelly, N. *Nature* **394**, 179–182 (1998).

Mathematics

Counting up to four

Ivar Ekeland

Place marbles in a straight line. You can always arrange them so that they are not evenly spaced. But can you arrange them so that no subset of k marbles, including non-neighbours, for $k \geq 3$, is evenly spaced (Fig. 1, overleaf)? Again, the answer is yes: if d is the distance between the first two marbles, just put the third one at distance $2d$ from the second one, the fourth at distance 2^2d from the third, and so on, the

n th being then at a distance $2^{n-2}d$ from the $(n-1)$ th. But then the distance between the marbles grows extremely fast, and they are spaced more and more widely apart: the density of the alignment (that is, the proportion of occupied sites between the first and the last one) tends to zero as the number of marbles increases. What happens if we impose non-zero density? The answer is that for any integer k and non-zero density δ there will be

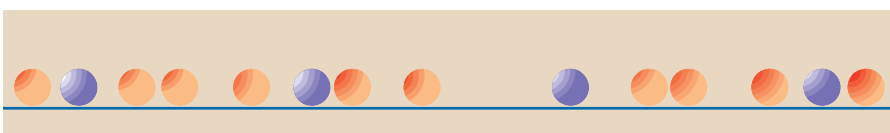


Figure 1 A game of marbles. These marbles may look randomly distributed, but four of them (in blue) are in fact evenly spaced. A calculation of the maximum number of marbles that can be placed in a line without any four of them being evenly spaced (or reaching zero density), has just been published⁵.

an upper limit $N_k(\delta)$ to the number of marbles that can be placed in a line, without any k of them being evenly spaced, and without the density of the alignment falling below δ .

In 1936, Erdős and Turán¹ stated this result as a conjecture — that is, something they believed was true but had no proof for. Some 40 years later it was proved for the first time by Szemerédi². In the meantime, Roth³ had given a very interesting proof for the case $k = 3$. Roth's proof is remarkable in that it tells us the largest possible alignment, of given density δ , that does not contain three evenly spaced balls: the maximum number of marbles in such an alignment turns out to be $\exp\{\exp(C/\delta)\}$, where C is a constant that does not depend on δ .

Now, a double exponential is certainly a very big number, but it pales in comparison with the much larger numbers that we get from Szemerédi's proof, and from subsequent ones, such as the remarkable proof by Furstenberg⁴, which relies on ergodic theory. Indeed, Roth's result was considered a notable achievement, and similar answers have long been sought for the case $k > 3$, in the hope that new estimates would lead to a better understanding of the combinations of numbers. This has just been achieved by Tim Gowers⁵, who proves that any alignment with density δ that does not contain four evenly spaced marbles must have fewer than $\exp\{\exp(1/\delta^C)\}$ marbles, where C is a constant which does not depend on δ . This is a breakthrough, because Gowers's method does not seem to be limited to $k = 4$; in fact, an unpublished manuscript by Gowers extends his result to all values of k , and gives a value of $\exp\{\exp(\delta^{-\exp\{\exp(k+10)\}})\}$ as an upper bound for the number of marbles in an alignment with density δ , which does not contain k evenly spaced marbles.

Of course, mathematicians do not like being caught playing with marbles, so they phrase the problem in terms of arithmetic progressions. Start with an integer n , pick another integer r , and count in succession $n + r, n + 2r, n + 3r$, and so on, up to $n + kr$: this is an arithmetic progression of length k . For instance, all multiples of 57 lying between 10 and 500 form an arithmetic progression of length eight. In this framework, Gowers's result can be restated as follows. Let N be a natural number greater than $\exp\{\exp(1/\delta^C)\}$, and consider the set $\{0, 1, \dots, N - 1\}$ of all natural numbers up to N , then every subset of size at least δN contains an arithmetic progression of length four.

Gowers's method builds upon Roth's inspiration, which was to use Fourier analysis. Roth shows that a subset A of $\{0, 1, \dots, N - 1\}$ will contain many arithmetic progressions of length three unless some Fourier coefficient is appropriately large with respect to the others, in which case a simple argument yields the desired estimate. If $k > 4$, the connection between the relative sizes of the Fourier coefficients and the presence of arithmetic progressions of length k is much weaker, and this is where previous attempts at extension broke down. This is not the place to follow Gowers's argument, except to say that at some point it relies on a beautiful theorem of Freiman⁶. This relates the structure of a set of natural numbers A , provided

we can estimate the size of $A + A$ (which is the set of all numbers $a + b$, where a and b belong to A), the first self-contained proof of which was provided by Ruzsa⁷.

Gowers's result raises hopes that one of the most famous problems in number theory might finally be solved: does the set of prime numbers contain arithmetic progressions of any length k ? In other words, given a number k , can one find an arithmetic sequence of length k consisting only of prime numbers? This question was raised by Erdős after he wrote his 1936 paper with Turán, and has remained unanswered ever since. □

Ivar Ekeland is at the University of Paris-Dauphine, 75775 Paris Cedex 16, France.

e-mail: ivar.ekeland@dauphine.fr

1. Erdős, P. & Turán, G. *J. Lond. Math. Soc.* **11**, 261–264 (1936).
2. Szemerédi, E. *Acta Arith.* **27**, 299–345 (1975).
3. Roth, K. F. *J. Lond. Math. Soc.* **28**, 245–252 (1953).
4. Furstenberg, H. *J. Anal. Math.* **31**, 204–256 (1977).
5. Gowers, W. T. *Geom. Funct. Anal.* **8**, 529–551 (1998).
6. Freiman, G. R. (in Russian) *Foundations of a Structural Theory of Set Addition* (Kazan Gos. Ped. Inst., Kazan, 1966). Translated in *Math. Monogr.* **37** (Am. Math. Soc., Providence, Rhode Island).
7. Ruzsa, I. *Acta Math. Hungarica* **65**, 379–388 (1994).

Volcanology

Fragmenting magma

Oleg E. Melnik

Explosive volcanic eruptions are among the most devastating of natural phenomena. The eruption of Santorini in the Bronze Age ended Minoan civilization; that of Vesuvius in AD 79 destroyed Pompeii; and that of Tambora, Indonesia, in 1815 killed 92,000 people. The hazards are by no means reduced these days, for as world population grows more people tend to live in areas that are potentially under threat. Hence the need to understand volcano behaviour, and whether eruptions are likely to take the explosive or gentler effusive form, and the motivation behind the work by Papale described on page 425 of this issue¹.

Qualitatively, explosive eruptions can be explained as follows. Magma with gas dissolved in it exists in a chamber in the Earth's crust, the chamber being connected with the surface by a narrow conduit. As magma ascends, pressure in it drops, the gas exsolves from the melt, and bubbles appear and grow in size. If conditions become appropriate for fragmentation of the magma, this bubbly flow changes into a gas–particle dispersion flow, and thus explosive eruption occurs.

It is this process that Papale has looked at anew. He proposes a strain-induced 'brittle fracture' mechanism for magma fragmentation, and calculates the conduit flow parameters for different types of magma using this principle. The fragmentation zone separates dense and viscous bubbly liquid below from the dilute gas–particle disper-

sion above. Thus the average mixture weight, conduit resistance and, therefore, the discharge rate of eruption, strongly depend on the position of the fragmentation level.

Study of explosive eruptions is difficult. They are relatively rare events, and in themselves highly dangerous, and often occur in inaccessible parts of the world. The governing parameters are known only approximately, and the characteristics of magma flow within the conduit cannot be reconstructed from indirect field measurements. Nor can the results of laboratory experiments easily be scaled-up to the natural situation because of the specific properties of magma. So we have to resort to computer simulations of eruptions to help understand their physics and forecast hazards for real volcanoes.

Simulations of conduit flow began in the late 1970s. By that time, physical models for bubbly liquid and gas–particle dispersion had become quite sophisticated but no fragmentation model for high-viscosity liquids had been put forward. In the case of vapour–water systems, which have been especially well investigated, the transition from a bubbly regime to a gas–droplet flow occurs when the bubbles coalesce into large packets. But the high viscosity of magmas prevents such coalescence and so the fragmentation mechanism must be different.

Several workers^{2–4} have used an assump-

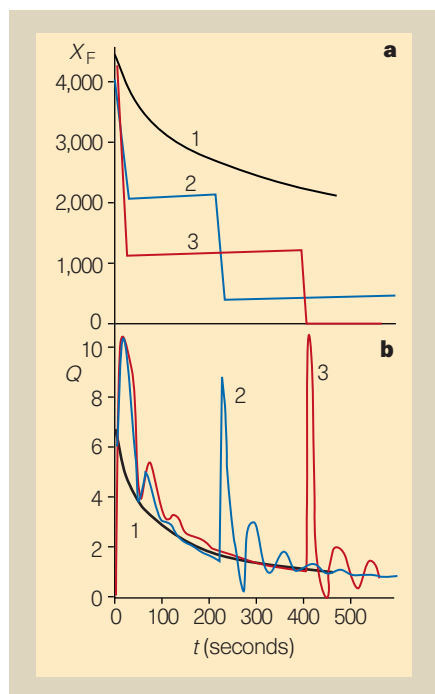


Figure 1 Simulation of the conduit flow dynamics for an explosive eruption of the Soufrière Hills volcano. a, Position of the fragmentation wave X_F (in metres, the magma chamber corresponds to $X_F = 0$) against time. b, Mass discharge rate of magma (Q , in 10^6 kg s^{-1}). Curves 1–3 correspond to three proposed mechanisms of magma fragmentation: fixed-volume concentration, strain induced and bubble overpressure. The calculations come from O. Melnik and R. S. J. Sparks, manuscript in preparation.

tion that the determining condition for magma fragmentation is when the growing bubbles reach their closest packing state; this is the fixed-volume concentration mechanism. A powerful argument against this assumption is that pumice shows a wide range in porosity. Typically the measured values are in the range from 0.5 to 0.9, and they vary from eruption to eruption or even during an eruption. Slezin⁵ suggested a more complicated mechanism for fragmentation, one which involved 'partly breaking foam', but this idea was not fully worked out mathematically.

Subsequently, Alidibirov⁶ and then Barmin and myself⁷ developed a shock-wave approach to the fragmentation of extremely viscous liquids. Alidibirov's theory describes the disruption of solidified magma with pressurized gas bubbles. Barmin and I assume that, in ascending magma, overpressure in growing bubbles develops because of the viscous resistance of surrounding magma, and this overpressure is taken to be the cause for fragmentation.

Now we have Papale's model¹, which goes back to the mid-nineteenth century and the ideas of James Clerk Maxwell on pseudo-fluids. According to Maxwell's equations,

when elongational strain rate in a mixture exceeds a critical value, determined by the mixture's physical properties, fragmentation occurs. Papale takes the critical strain-rate value from Alidibirov and Dingwell's experiments⁸ on shock magma fragmentation. This approach leads to the same mathematical formulation as that which we proposed⁷, and the same relation between the porosity of magma fragments and magma viscosity (the more viscous the magma, the denser the fragments).

The influence of the various proposed fragmentation mechanisms can be illustrated as they bear upon modelling of the opening of a volcanic conduit covered by a solidified magma plug. This situation is a common one at the beginning of an explosive eruption. Figure 1 shows the position of the magma fragmentation zone and the magma discharge rate against time for a set of parameters pertaining to the 17 September 1997 eruption of the Soufrière Hills volcano on Montserrat. For the fixed-volume concentration mechanism (curve 1), discharge rate decreases monotonically and the system reaches a steady state. By contrast, application of the strain-induced and bubble-overpressure approaches (curves 2 and 3) gives an eruption with several pulses, each pulse representing a fragmentation event, separated by periods when bubbly liquid at depth ascends without fragmentation. These pulses are thought to correspond to those observed seismically during the eruption. Curve 3 is in better agreement with the seismic data, but it still gives a shorter interval between the calculated pulses than those that were measured.

The problem of magma fragmentation is far from solved, however. Future research must be aimed at creating a more sophisticated model, which accounts for the microphysics of bubble growth and magma rheology. Unfortunately, further progress will be impossible unless there is a great improvement in our understanding of the mechanics of multiphase systems on both small and large scales. Nonetheless, both strain-induced and bubble-overpressure mechanisms seem to be key factors in magma fragmentation and thus explosive eruptions. □

Oleg E. Melnik is at the Institute of Mechanics, Moscow State University, 117192, 1-Mitshurinski Prospect, Moscow, Russia.
e-mail: melnik@inmech.msu.su

- Papale, P. *Nature* **397**, 425–428 (1999).
- Wilson, L., Sparks, R. S. J. & Walker, G. P. L. *Geophys. J. R. Astron. Soc.* **63**, 117–148 (1980).
- Wilson, L. *J. Volcanol. Geotherm. Res.* **8**, 297–313 (1980).
- Dobran, F. *J. Volcanol. Geotherm. Res.* **49**, 285–311 (1992).
- Slezin, Yu. B. *Vulkanologiya i Seismologiya* **N1**, 35–45 (1991).
- Alidibirov, M. A. *Bull. Vulkanol.* **56**, 459–465 (1994).
- Barmin, A. A. & Melnik, O. E. *Izv. Ros. Akad. Nauk. Mekh. Zhidk. Gaza* **2**, 49–60 (1993). Translated in *Fluid Dynamics* **28**, 195–202 (1993).
- Alidibirov, M. A. & Dingwell, D. B. *Nature* **380**, 146–148 (1996).

Daedalus

Powerful lighting

Flames and explosions propagate by the movement of molecules, either as reactive radicals or a dense shock-pulse.

Accordingly, even the highest explosives detonate at 'speed of sound' velocities — a few kilometres per second. By contrast, an explosion which propagated by radiation should detonate at near the speed of light. Daedalus is now trying it.

His first idea was to silver a long, thin crystal of a solid high explosive, or fill a silvered tube with a liquid one, to create a sort of explosive light-pipe. The silvering would guide the radiation of explosion along the pipe, igniting the charge as it went. Sadly, the radiation would probably be too weak and diffusely absorbed to do the job, especially if it had to set off the explosive by bulk heating. But Daedalus recalls the photosensitizing dyes used in photographic emulsions. They absorb a wavelength to which the emulsion would otherwise be blind, and capture its energy for photochemical action. So DREADCO chemists are seeking dyes to do the same thing for nitroglycerine and TNT. The dye must absorb radiation of a wavelength emitted copiously by the explosion, and use it to drive the very first step of the explosive reaction. Once the light energy has excited this 'uphill' activation step, the reaction will then run vigorously and spontaneously 'downhill', pouring out radiation to set off molecules nearby. The dye molecules should ideally be bonded to those of the explosive, but may only need to be included in its crystal lattice.

Photosensitized explosive will be tricky to handle. With luck its triggering wavelength will be somewhere in the near UV, only weakly present in daylight. Even so, just as one neutron can set off an assembled nuclear charge, so one photon will detonate any amount of photoexplosive if it has a compact optical layout with a photon-gain greater than unity. The product will have to be stored in thin sheets between black paper, and only assembled in its silvering just before use.

But it will then transform blasting and explosive forming. It will deliver energy to rock or metal far faster than it can be dissipated by shock waves or any other molecular motion. Even better, unlike thermally detonated charges, photon-fired ones will work on the smallest of scales. Fired by precision-guided laser pulses, photoexplosives will find their way into small-scale technologies from engraving, dentistry and micro-surgery right down to integrated circuit manufacture.

David Jones

Obituary

Hans Oeschger (1927–98)

Pioneer in environmental physics

The research of Hans Oeschger, who died on 25 December 1998 after a long illness, centred on applying the methods of modern physics to investigations of the Earth system. In particular, he and his team in Bern produced a stream of results that have become cornerstones of studies on climate change.

Oeschger was trained as an experimental physicist at ETH Zürich, and learned to appreciate the simplicity of physical laws. But he also recognized the limitations of the classical method in trying to dissect a system into its components, whereby processes fundamental to understanding the system in its entirety are lost. In F. G. Houtermans in Bern he found the right mentor for his dissertation: “Oeschger wanted to become a musician. However, since I do physics much like an artist, he now feels quite comfortable in physics”.

The goal of Oeschger’s dissertation was to build a β -counter to measure ultra-low radioactivity by counting the electrons emitted during decay. In 1955 he completed this device, now known as the Oeschger counter. It had a lower background than any other available instrument, due to a clever anti-coincidence technique which dismissed counts that were not caused by the decaying sample, and became for many years the leading instrument in radiocarbon laboratories. In 1959, using this device, Oeschger, Suess and Rakestraw were the first to succeed in radiocarbon-dating Pacific deep water. They went on to measure the activity of a variety of naturally occurring radioisotopes, and embarked on a long journey to quantify the numerous exchange processes between different components of the Earth system.

In 1962 Oeschger started to determine the tritium contents of firn (unconsolidated snow) and ice, for use as archives of environmental information. His involvement in expeditions to Greenland and Antarctica between 1964 and 1992 gave his laboratory direct access to polar ice cores. The analytical methods and the drilling techniques that Oeschger and his team devised enabled them to produce an unprecedented reconstruction of climate changes of the past 150,000 years.

Oeschger was a leading figure in ice-core research and one of the fathers of the deep-drilling projects Dye 3 and GRIP in



Greenland. Along with his colleagues he was the first to measure the glacial–interglacial change in levels of atmospheric CO₂. In 1979 they showed that the atmospheric concentration of CO₂ during the last glacial period was almost 50% lower than today. Furthermore, they analysed the CO₂ changes during the past 1,000 years and showed that the dramatic increase in industrial times is a direct consequence of the burning of fossil fuels.

Hans Oeschger was troubled by the potentially adverse consequences of an increased greenhouse effect caused by the steady rise in atmospheric CO₂. His early insight was based on a novel geochemical box-diffusion model for the global carbon cycle that his team developed. He also took his wider responsibilities as a scientist very seriously: “The worst for me,” he declared, “would be if there were serious changes in the next five to ten years and we scientists ... did not have the courage to point out these dangerous developments early on”. He gave many public lectures on this and other topics, and was also a lead author of the First Assessment Report of the Intergovernmental Panel on Climate Change.

There were even more surprises to come from the Greenland ice cores. Together with Chester C. Langway in the United States and Willi Dansgaard in Denmark, Oeschger documented a series of abrupt climate changes in the cores. Stable isotope measurements on carbonate sediments from Lake Gerzensee, near Bern, showed that these changes were not only local to Greenland but reflected large climatic swings of at least hemispheric extent. Twenty-four such swings, termed Dansgaard–Oeschger events, are evident during the last glacial in many different palaeoclimatic archives.

As long ago as 1984, Oeschger

recognized the importance of ocean circulation in explaining abrupt climate change and used the physical analogy of a flip-flop system — triggered by small perturbations, the ocean circulation might switch from one circulation mode to another. Based on these insights, he was among the first to point out that the anthropogenic increase of CO₂ could represent such a perturbation. Although his early warnings were often greeted with disbelief, the latest results from both observations and high-resolution palaeoclimate archives are an impressive testimony to his forward thinking on global warming.

One of Oeschger’s great strengths was his recognition that the Earth must be seen as a complex system. Long before inter- and transdisciplinarity became buzz-words, his physical intuition and readiness to listen allowed him to weave links between seemingly disconnected observations and concepts. Hans’s personal warmth and enthusiasm led to many fruitful collaborations across disciplinary boundaries. But he was always firm in pointing out that successful transdisciplinary research needs input of the highest quality from the disciplines concerned. The wide respect in which he was held stemmed, not least, from his own fundamental contributions to physics.

We owe much to Hans Oeschger. He would disguise his often challenging suggestions with the gentle and informal Swiss-German phrase, “*Das sötted mir doch chöne mässe*” (“We ought to be able to measure this”), words that on many occasions marked the beginning of a novel method to quantify environmental processes. With this approach, his contributions to palaeoclimatology transformed a descriptive field into a science in which changes of environmental conditions can be given numbers and units. His results and way of thinking will guide us on our way for a long time yet. Moreover he saw his science as being of wider service: “Real progress in this field will make it possible for society to act based on foresight. The anticipated climate change is only one of the great problems with which society is faced; the thorough way climate change is assessed by the scientific community may serve as an example for addressing other grand challenges”. Those privileged to have worked or shared ideas with him will not forget his friendship.

Thomas Stocker

Thomas Stocker is in Climate and Environmental Physics, Physics Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland. e-mail: stocker@climate.unibe.ch

BUCHER, BERN

Half-awake to the risk of predation

Birds have overcome the problem of sleeping in risky situations by developing the ability to sleep with one eye open and one hemisphere of the brain awake¹. Such unihemispheric slow-wave sleep is in direct contrast to the typical situation in which sleep and wakefulness are mutually exclusive states of the whole brain. We have found that birds can detect approaching predators during unihemispheric slow-wave sleep, and that they can increase their use of unihemispheric sleep as the risk of predation increases. We believe this is the first evidence for an animal behaviourally controlling sleep and wakefulness simultaneously in different regions of the brain.

The function of unihemispheric slow-wave sleep (USWS) in birds has been unclear¹. In aquatic mammals — the only other group of animals known to exhibit USWS — this sleep pattern seems to allow concurrent sleep and surfacing to breathe². If avian USWS functions as a form of predator detection, then birds should be able to control whether they sleep primarily with one or both hemispheres in response to changes in predation risk³. Under safe conditions, birds can sleep more efficiently by sleeping with both hemispheres simultaneously, whereas under dangerous conditions a bird should increase USWS to remain vigilant for predators. Moreover, during USWS a bird should sleep with its open eye towards the direction from which a potential predator is likely to approach.

To test these expectations, we took advantage of the 'group edge effect', a phenomenon in which animals at the more risky edge of a group spend more time scanning visually for predators than those nearer the group's centre⁴. We predicted that birds sleeping at the edge of a group would spend more time in USWS, and would direct their open eye away from the group's centre towards potential predators.

We established four groups of mallard ducks (*Anas platyrhynchos*), each containing four individuals. Video recordings of sleep behaviour were made with the birds arranged in a row. Birds at the ends of the row (the edge of the group) were in a more exposed position than those in the central positions. The closure of one eye indicated USWS¹, whereas closing both eyes indicated either bihemispheric slow-wave sleep or rapid-eye-movement (REM) sleep, which occurs bihemispherically in birds⁵.

We found that mallards markedly increased their use of USWS when sleeping at the edge of the group. The proportion of sleep composed of USWS increased from $12.4 \pm 1.1\%$ (mean $\pm$ s.e.) in the central position to $31.8 \pm 3.6\%$ in the edge position

(Wilcoxon matched-pairs signed-ranks test, $T = 1$, $P < 0.001$), an increase of more than 150%. Furthermore, birds in USWS at the edge position oriented the open eye away from the group's centre $86.2 \pm 3.1\%$ of the time (t -test, $t = 11.58$, d.f. = 15, $P < 0.001$), whereas birds in the central position showed no preference for gaze direction (time looking away, $52.8 \pm 3.7\%$; t -test, $t = 0.76$, d.f. = 15, n.s.). Mallards therefore exhibited behavioural control of USWS and used it adaptively by looking in the direction of a potential threat.

Electroencephalographic (EEG) recordings obtained under the above conditions

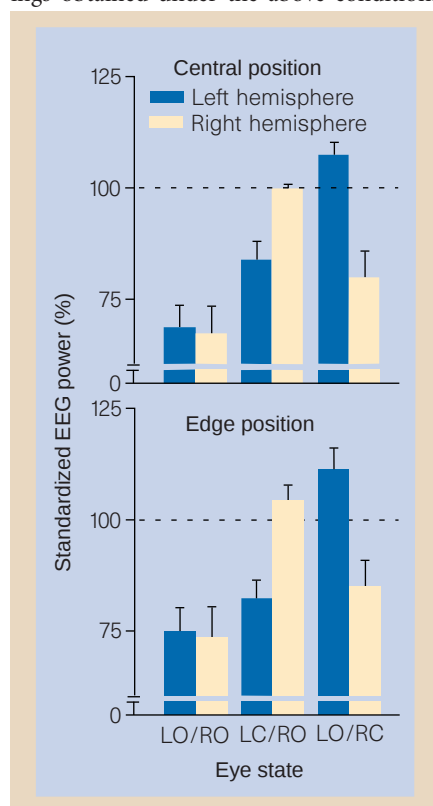


Figure 1 The relation between eye state and standardized EEG power (1–6 Hz) for the left and right hemispheres of birds occupying central and edge positions. EEGs were recorded from the dorsal surface of the hyperstriatum accessorium of each hemisphere, the part of the avian brain that most clearly reflects sleep-related changes in brain state⁶. The three eye states are: both left and right eyes open (LO/RO); left eye closed and right eye open (LC/RO); and left eye open and right eye closed (LO/RC). For each mallard ($n = 6$), EEG power was standardized as a percentage of the average power observed during bihemispheric slow-wave sleep for each bird's hemisphere, so the broken line at 100% indicates EEG power equivalent to that during bihemispheric slow-wave sleep. Greater power in the hemisphere opposite the closed eye during unilateral eye closure indicates USWS. Further methodological details are available from the authors.

corroborated at the neurophysiological level our behavioural results. During each eye state, EEG recordings from each hemisphere were evaluated by digital period amplitude analysis, which quantifies the power (a measure of wave amplitude) of different frequencies within the EEG⁶. Avian slow-wave sleep is characterized by high power in the low-frequency range (1–6 Hz), whereas wakefulness is characterized by low power in this range⁵.

When one eye was closed, low-frequency power in the hemisphere opposite the closed eye (the sleeping hemisphere) was significantly greater than that in the hemisphere opposite the open eye ($T \leq 1$, $P < 0.05$ for all comparisons; Fig. 1), regardless of the position occupied by a bird, which confirms USWS. Low-frequency power in the awake hemisphere (that opposite the open eye) was nevertheless significantly greater than the power with both eyes open ($T = 0$, $P < 0.05$ for all comparisons), indicating a quiet waking state intermediate between full alertness and slow-wave sleep. Subsequent tests confirmed that the waking hemisphere was capable of predator detection: mallards in USWS initiated escape behaviour within only 0.165 ± 0.006 seconds ($n = 10$ mallards) of an expanding video image being presented, simulating predatory attack.

The EEG recordings confirmed that the 150% increase in unilateral eye closure was a direct result of an equivalent increase in the proportion of slow-wave sleep composed of USWS. Because bihemispheric slow-wave sleep and REM sleep both occur with both eyes closed, it was possible that a selective decrease in REM sleep in the edge position may have confounded our interpretation of the behavioural results. However, EEG analyses showed an almost 1:1 relation between unilateral eye closure and the proportion of slow-wave sleep composed of USWS. An increase in unilateral eye closure therefore reflected an equivalent increase in USWS.

A bird's ability to control sleep and wakefulness independently in each hemisphere is in accord with evidence indicating that each eye⁷ and associated hemisphere⁸ can function independently during wakefulness in birds. The neuroanatomical structures and physiological processes that allow independent hemispheric functioning during wakefulness may therefore also allow for the independent hemispherical control of sleep. The ability to control USWS behaviourally in response to changes in predation risk emphasizes the dynamic balance that animals must achieve between the ecological consequences and the physiological necessity of sleep⁹.

Niels C. Rattenborg, Steven L. Lima,
Charles J. Amlaner

Department of Life Sciences,
Indiana State University,
Terre Haute, Indiana 47809, USA
e-mail: lsratten@scifac.indstate.edu

1. Ball, N. J., Amlaner, C. J., Shaffery, J. P. & Opp, M. R. in *Sleep* '86 (eds Koella, W. P., Obál, F., Schulz, H. & Visser, P.) 151–153 (Fischer, New York, 1988).
2. Oleksenko, A. I., Mukhametov, L. M., Polyakova, I. G., Supin, A. Y. & Kovalzon, V. M. *J. Sleep Res.* **1**, 40–44 (1992).
3. Lima, S. L. *Adv. Study Behav.* **27**, 215–290 (1998).
4. Elgar, M. A. *Biol. Rev.* **64**, 13–33 (1989).
5. Amlaner, C. J. & Ball, N. J. in *Principles and Practice of Sleep Medicine* 2nd edn (eds Kryger, M. H., Roth, T. & Dement, W. C.) 81–94 (Saunders, Philadelphia, 1994).
6. Hoffmann, R. F., Moffitt, A. R., Shearer, J. C., Sussman, P. S. & Wells, R. B. *Waking Sleeping* **3**, 1–16 (1979).
7. Schaeffel, F., Howland, H. C. & Farkas, L. *Vision Res.* **26**, 1977–1993 (1986).
8. Bredenkötter, M. & Bischof, H.-J. *Vis. Neurosci.* **5**, 155–163 (1990).
9. Rechtschaffen, A., Gilliland, M. A., Bergmann, B. M. & Winter, J. B. *Science* **221**, 182–184 (1983).

Pattern formation in semiconductors

In semiconductors, nonlinear generation and recombination processes of free carriers and nonlinear charge transport can give rise to non-equilibrium phase transitions^{1,2}. At low temperatures, the basic nonlinearity is due to the autocatalytic generation of free carriers by impact ionization of shallow impurities. The electric field accelerates free electrons, causing an abrupt increase in free carrier density at a critical electric field. In static electric fields, this nonlinearity is known to yield complex filamentary current patterns bound to electric contacts³.

We used microwaves to apply an electric field to semiconductor samples without using electrical contacts. High-frequency electric fields ionize impurities just as d.c. fields do⁴, but they do not impose inhomogeneities like electric contacts. We find that in thin n-type gallium arsenide (GaAs) epitaxial layers subjected to a uniform microwave field, circular spots of enhanced free electron density with sharp boundaries are spontaneously formed above a critical microwave threshold power. This new type of self-organized free-carrier density pattern is different from current filaments in that they are currentless; they also differ from electron-hole drops⁵ as only one type of charge carrier is involved.

The spatial patterns in free electron density were made visible by photoluminescence quenching⁶. Samples cooled to low temperatures (1.8 K) were illuminated by interband light and photographed in the spectral range of the luminescence of exciton recombination and donor–acceptor transitions. With increasing microwave power P , a decrease in photoluminescence occurs at a threshold value P_+ as a result of

an almost circular spot of enhanced electron density with a diameter (D_+) of about 1 mm (Fig. 1b). The diameter of the spot at the threshold is always finite and independent of the size of the semiconductor sample. This pattern formation was observed in doped samples with impurity densities of about 10^{15} per cm^3 , but not in 'ultrapure' material with an impurity density of about 5×10^{12} per cm^3 .

When the microwave power is increased above the threshold P_+ , the diameter of the original spot increases (Fig. 1b,c); at certain values of P , additional spots appear at a distance of 1 to 3 mm (Fig. 1d). Decreasing the microwave power after the first spot has formed makes it smaller until it vanishes at a power of $P_- < P_+$, at a diameter $D_- < D_+$ (Fig. 1e). The pattern formation process shows a hysteretic behaviour, as it is characteristic for a first-order phase transition.

The quenching of the photoluminescence in the spots indicates that the average energy of electrons is high enough to ionize impurities and excitons in the spots. The observed structures therefore correspond to spots of high free electron density. We verified this using a sample with two parallel stripe contacts at opposite edges. When a voltage was applied across the contacts, a current filament was formed in addition to the microwave-induced spot. The current filament goes through the spot, indicating that the observed structures are regions of high free electron density.

Our findings indicate that the physical background of microwave-induced pattern formation in the electron density is the

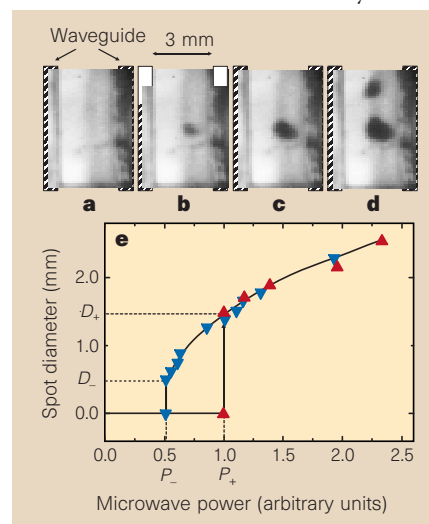


Figure 1 Self-organized formation of spots of high electron density in an n-doped gallium arsenide epitaxial layer as a result of impact ionization of shallow donors in a microwave electric field. **a–d**, Images of near-infrared luminescence for microwave power increasing from zero (**a**) to about 30 mW (**d**). Shaded stripes show the walls of a waveguide. **e**, Spot diameter as a function of microwave power. Red and blue triangles indicate increasing and decreasing power, respectively. P_+ is of the order of 10 mW.

same as that of current filamentation. Both of these phenomena are based on a bistability of conductivity combined with a constraint of the external driving mechanism. In the case of current filaments, the constraint is provided by the finite current applied to the sample, whereas for microwave-induced pattern formation the external constraint must be the limited power supply from the field.

Bistability and the possibility of spatial modulation of semiconductor conductivity have been attributed to three different physical mechanisms: multilevel generation and recombination kinetics of impurities², run-away of electron energy due to energy-dependent electron scattering⁷, and electron density-dependent screening of ionized impurity scattering⁴. For all three mechanisms, bistability vanishes below a critical density of impurities, which explains the qualitatively different behaviour of doped and ultrapure materials, for which pattern formation has not been observed.

V. V. Bel'kov*, J. Hirschinger*, V. Novák†, F.-J. Niedernostheide*, S.D. Ganichev*, W. Prettl*

*Institut für Experimentelle und Angewandte Physik, Universität Regensburg, 93040 Regensburg, Germany

†Institute of Electrical Engineering, AV ČR, 182 02 Prague 8, Czech Republic
e-mail: wilhelm.prettl@physik.uni-regensburg.de

1. Landsberg, P. T. & Pimpale, A. J. *Phys. C* **9**, 1243–1252 (1976).
2. Schöll, E. *Nonequilibrium Phase Transitions in Semiconductors* (Springer, Berlin, 1987).
3. Ridley, B. K. *Proc. Phys. Soc.* **82**, 954–966 (1963).
4. Kozhevnikov, M. et al. *Phys. Rev. B* **52**, 4855–4863 (1995).
5. Jeffries, C. D. *Science* **189**, 955–964 (1975).
6. Eberle, W. et al. *Appl. Phys. Lett.* **68**, 3329–3331 (1996).
7. Levinson, I. B. *Sov. Phys. Solid State* **7**, 1098–1102 (1965).

UV-B damage amplified by transposons in maize

While absorbing visible light energy for photosynthesis, plants are unavoidably exposed to ultraviolet radiation, which is particularly harmful at shorter wavelengths (UV-B radiation). Ozone depletion in the atmosphere means that plants receive episodic or steadily increasing doses of UV-B, which damages their photosynthetic reaction centres, crosslinks cellular proteins, and induces mutagenic DNA lesions¹. Plant adaptive mechanisms of shielding and repair are therefore critical to survival — for example, somatic tissues of maize and *Arabidopsis* defective in phenolic sunscreen pigments^{2,3} incur increased DNA damage, and mutants defective in DNA repair^{4,5} are killed by UV-B.

The harmful effects of UV-B on maize pollen are proportional to the exposure time⁶. I find that simulated field conditions

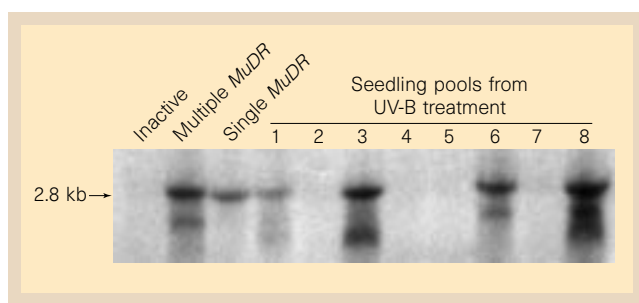
of UV-B exposure at 33% ozone depletion^{1,6} can activate immobile *Mutator* (*Mu*) transposons⁷ in maize sperm. These transposons amplify the effects of UV-B exposure by causing mutations beyond the extent of immediate DNA damage.

Mu elements, defined by shared terminal inverted repeat (TIR) ends, are regulated by *MuDR*, which encodes MURA transposase, which binds to the *Mu* TIRs⁸ (Fig. 1a). When *MuDR* is transcriptionally active, *Mu* elements insert (causing mutations) and excise from existing locations (creating phenotypic variegation, a property seen as somatic spotting in pigment genes; Fig. 1b). When Mutator is inactive, *Mu* elements are methylated⁷, *MuDR* transcripts are undetectable (Fig. 2), and no excision occurs (Fig. 1c). Spontaneous reactivation of somatic excision is rare (frequency $< 10^{-4}$). First-generation (F_1) inactive Mutator plants, homozygous for *bz2::Mu1* or *bz2::MuDR* reporter alleles in the anthocyanin pigment pathway, reactivate when crossed to an active Mutator *bz2* line (6 of 25 and 5 of 15 progeny ears had spotted kernels; see Supplementary Information). Reactivation crosses are more efficient with a maternal active Mutator parent (11.5% spotted kernels compared with paternal 1.9% spotted kernels). Reactivation frequency falls with each generation in which an inactive Mutator line is maintained. After three generations, reactivation occurs in fewer (3 of 29) cases and less efficiently (0.05% spotted kernels).

Irradiation of pollen with UV-B for 3 minutes restored *Mu* excision in 7 of 19 progeny ears (6.2% spotted kernels) in F_1 inactive plants. The population has considerable heterogeneity in reactivation potential as both reactivation crosses and pollen irradiation yielded progeny with a wide range of degree of spotting. Reactivation was much lower in F_3 inactive individuals, but UV-B was 14-fold more effective than crossing to a *bz2* active Mutator plant.

Maize pollen contains three haploid cells: a vegetative cell and two sperm. One sperm fertilizes the egg to form the diploid zygote, whereas the other contributes to the

Figure 2 Northern blot to detect *mudrA* transposase transcripts. Total RNA samples were from pools of three seedlings; blots were hybridized with a *mudrA* transposase-specific probe (Fig. 1a). The 2.8-kb transcript encodes an 823-amino-acid transposase⁸ required for Mutator activity; active lines produce other smaller transcripts from alternative splicing of the transposase mRNA and from internally deleted *MuDR* elements.



formation of endosperm, the tissue where somatic excision is scored. If UV-B reactivates Mutator activity in individual nuclei, usually one sperm will be reactivated (non-concordance); but if UV-B physiologically changes pollen to activate Mutator secondarily, both sperm could be affected (concordance). I therefore checked whether kernels with spotted endosperm produced a plant that transmitted reactivated Mutator to the next generation and found that UV-B induced 19 spotted:1,020 bronze kernels; none of the 19 had progeny with spotted kernels. Of 350 plants grown from bronze siblings, nine crosses to *bz2* and 14 self-pollinations produced spotted progeny, so sperm non-concordance normally occurs.

Transcriptional activation of silent *MuDR* elements could be a nucleus-specific mechanism underlying non-concordance. To test this idea, 24 bronze progeny were germinated from an ear with 23% spotted endosperms. If embryo reactivation is similar, 54% of pools from three seedlings should have at least one individual with a reactivated Mutator system: $100(1 - (0.77)^3)$. Of eight pools of bronze seedlings, four contained *mudrA* transcripts (Fig. 2), indicating that seedling transcriptional reactivation parallels visible endosperm spotting.

Doubling terrestrial UV-B (> 290 nm) is cytotoxic to pollen, increasing the frequency of new mutations⁶ but not by enough to impair agriculture. Activation of cryptic transposable elements, however, could increase the mutation rate. Although

most transposons are transcriptionally silent and immobile, they can cause many spontaneous mutations in maize and other organisms⁹. Transposons may be activated during 'genomic shock' as an adaptive mechanism¹⁰. Whereas damage to DNA is immediately repaired or fixed as stable mutations, transposons produce cycles of insertion and excision long after activation.

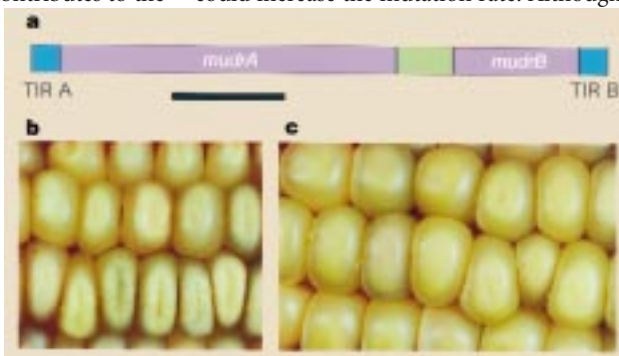
Plants with large genomes may be particularly susceptible to destabilization. Inactive retrotransposons constitute about half of the maize genome¹¹ and so are a reservoir of potential mutagens. Haploid pollen is a sensitive target of UV-B irradiation. Because plants lack a dedicated germ line, mutations in somatic stem cells can be represented in subsequent gametes¹². Normally stringent selection on the vegetative cell of haploid pollen, which directs pollen maturation and growth, counterbalances the accumulation of deleterious alleles in the diploid soma. Mutations induced in individual sperm are not subjected to such rigorous selection, because genetic activity is thought to start after fertilization. As a result, transposon activation in individual sperm could greatly increase the genetic load in higher plants.

Virginia Walbot

Department of Biological Sciences,
Stanford University,
Stanford, California 94305-5020, USA
e-mail: walbot@stanford.edu

Figure 1 Mutator transposon properties.

a, The 4.9-kilobase (kb) *MuDR* element showing the convergent transcription units of the *mudrA* transposase and helper function *mudrB* gene. Transcription initiates in the TIRs (blue) and terminates in the intergenic region (green); the *mudrA* hybridization probe region is underlined. **b**, Excision of *Mu1*,



a 1.4-kb non-autonomous element sharing only the TIRs with *MuDR*, from the *bz2::Mu1* allele in a UV-B-reactivated Mutator line restores *Bronze2* gene expression in spots of anthocyanin pigmentation. **c**, No excision from *bz2::Mu1* occurs in an inactive Mutator sibling. Details of maize lines and irradiation protocols are available from the author.

- Rozema, J., van de Staaij, J., Björn, L. O. & Caldwell, M. *Trends Ecol. Evol.* **12**, 22–28 (1997).
- Stapleton, A. E. & Walbot, V. *Plant Physiol.* **105**, 881–889 (1994).
- Landry, L. G., Chapple, C. C. S. & Last, R. L. *Plant Physiol.* **109**, 1159–1166 (1995).
- Britt, A. B., Chen, J.-J., Wykoff, D. & Mitchell, D. *Science* **261**, 1571–1574 (1993).
- Landry, L. G. *et al. Proc. Natl Acad. Sci. USA* **94**, 328–332 (1997).
- Chrispeels, H. E. *Effects of Ultraviolet Radiation on Maize*. Thesis, Stanford Univ. (1996).
- Walbot, V. *Mol. Gen. Genet.* **234**, 353–360 (1992).
- Benito, M.-I. & Walbot, V. *Mol. Cell. Biol.* **17**, 5165–5175 (1997).
- Kidwell, M. G. & Lisch, D. *Proc. Natl Acad. Sci. USA* **94**, 7704–7711 (1997).
- McClintock, B. *Science* **226**, 792–801 (1984).
- SanMiguel, P. *et al. Science* **273**, 765–769 (1996).
- Walbot, V. *Trends Plant Sci.* **1**, 27–32 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Collective decisions and cognition in bees

In a remarkable example of collective decision-making, swarms of honeybees, *Apis mellifera*, choose one of many nest sites discovered and reported by their scouts. At first, dancing scouts communicate the location of many sites, but within a few days all dances focus on the same high-quality site^{1–3}. Instead of swarms acquiring global information by direct comparison of sites^{4,5}, we find that the swarm's decision arises through a self-organized process driven by the dynamics of interacting individuals following simple rules based on local information.

The number of scouts increases at nest sites through the positive-feedback process of recruitment, and decreases by attrition at less-favoured sites, allowing unanimity to develop⁶. Two hypotheses have been proposed to explain the attrition: scouts either compare sites directly and choose the better one⁷, or they drop out of the process after reporting on sites they find¹. All three scouts observed in one study¹ visited (and perhaps compared) more than one site. If such a comparison really is crucial to the decision-making process of bee swarms, it would indicate that individual bees are capable of complex cognitive processing^{7,8}.

We tested whether scouts that dance for one site are selective in dance-following. Do they follow dances for their own site (to assess the number of scouts investigating it), for sites they have not visited (such crossover dance-following would be indicative of cognitive ability⁷), or do they follow dances at random? In a desert area where nest sites could be controlled, we observed swarms while they chose between two identical nest boxes⁶. We marked each dancer

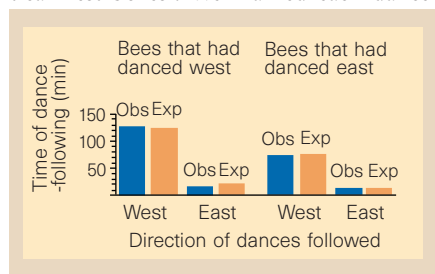
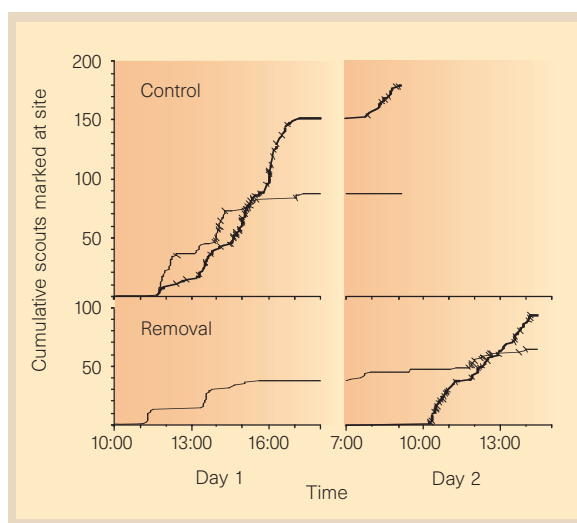


Figure 1 Time for which bees that had danced for the east or west site then followed dances for each site. Obs, observed, and Exp, expected, dancing times. For site S (and for site A), n_s is the number of bees that were marked dancing for site S, F_i is the total time bee i spent following dances, D_S and D_A are the number of dances for S and A (excluding those of bee i) during minute t_i and t_i is the time at which bee i first danced. Expected (S) in minutes is given by $\sum_{i=1}^{n_s} (F_i \times \sum_{t=t_i}^{t_{takeoff}} (D_S)_t / \sum_{t=t_i}^{t_{takeoff}} (D_S + D_A)_t)$. There is no significant difference between observed and expected values ($\chi^2 = 0.41$, $P = 0.94$).

Figure 2 Cumulative number of scout bees marked at south and north sites as the swarm decided between them. South, heavy line; north, fine line. Crossover occurred in both directions (average of $18 \pm 10\%$ of scouts; $n = 6$) and was the same for favoured and unfavoured sites, and 'removal' and control treatments (arcsine transform: $t_0 = 1.50$, $P > 0.19$; $t_4 = 1.06$, $P > 0.38$). All swarms reached unanimity (two controls chose north and one chose south; one 'removal' chose north and two chose south). In each treatment, two swarms moved on the second day after they were set up (as here), and one on the third day. Total scouting time until movement (including only time between first and last dancing each day) did not differ: 13.0 ± 3.3 hours in three 'removal' replicates, and 13.2 ± 1.7 hours in three control replicates ($t = 0.12$, $P > 0.9$).



individually when she first danced, and videotaped all dancing and dance-following by marked bees throughout the decision-making process. The dancing performed for each of the two sites was measured from the discovery of the first nest site to the time the swarm took off. We calculated the distribution of dances available for each marked dancer that followed her first dance, and found that dance-following fits this random expectation (Fig. 1): scouts followed dances for two sites in proportion to the total amount of dancing by others for these sites.

We then tested the role of direct comparison in the decision-making process. If comparison is prevented, the bees might fail to reach a unanimous decision, causing them to take longer to decide. We marked all scouts that inspected each site, and each bee that danced, with coloured paint. In 'removal' replicates, we captured and removed 'crossover' bees that visited a second site; in 'control' replicates, we captured, marked and released crossovers. Preventing bees from making comparisons and dancing for the alternative site did not prevent or delay swarms from arriving at a decision and taking off. The build-up of scouts at each nest box is shown in Fig. 2 for the shortest replicates: the time course of the decision is similar in these, as it is overall.

Direct comparison of potential nest sites by scouts is therefore probably not an important part of the collective decision-making process by swarms. Instead, the decision arises from local information and the properties of the underlying dynamic interactions among individuals, who do not need a global view, which is characteristic of self-organized systems^{4,5,9}. Recruitment's positive feedback (probably modulated with respect to site qualities⁶ such as nectar availability¹⁰) and stochastic factors, including individual differences between bees and time of discovery⁹, would account for

differences in build-up, but not for attrition of scouts from less-favoured sites.

As direct comparison and crossover to favoured sites do not drive attrition, it may be more important that bees drop out of the process after some dancing, leaving the process to succeeding cohorts of scouts^{1,6}. This is one way in which recruitment to nest sites differs from that to food sources, as it allows differences in positive feedback between sites to drive the process to unanimity by overwhelming the lower recruitment for less-favoured sites.

Thus the collective decision of a swarm seems to be based not on complex cognitive comparison, but on more limited cognitive tasks and information feedback. Although it is self-organized, this process has been shaped by natural selection on the behavioural components (such as scouts dropping out and recruiters following the dances of others) that determine its workings. Nature can build complexity in this way by using local information and simple rules.

P. Kirk Visscher*, Scott Camazine†

*Department of Entomology,
University of California,
Riverside, California 92521, USA
e-mail: visscher@citrus.ucr.edu

†Department of Entomology,
Pennsylvania State University,
State College, Pennsylvania 16802, USA

1. Lindauer, M. Z. *Vergl. Physiol.* **37**, 263–324 (1955).
2. Lindauer, M. *Nature* **179**, 63–66 (1957).
3. Lindauer, M. *Naturwissenschaften* **40**, 379–385 (1951).
4. Camazine, S. & Sneyd, J. J. *Theor. Biol.* **149**, 547–571 (1991).
5. Deneubourg, J.-L., Aron, S., Goss, S. & Pasteels, J. M. *J. Insect Behav.* **3**, 159–168 (1990).
6. Camazine, S., Visscher, P. K., Finley, J. & Vetter, R. S. *Insectes Soc.* (in the press).
7. Griffin, D. R. *Animal Minds* (Univ. Chicago Press, 1992).
8. Gould, J. L. & Gould, C. G. *The Animal Mind* (Freeman, New York, 1994).
9. Camazine, S. et al. *Self-organization in Biological Systems* (Princeton Univ. Press, NJ, in the press).
10. Seeley, T. D. *The Wisdom of the Hive: The Social Physiology of Honey Bee Colonies* (Harvard Univ. Press, Cambridge, MA, 1995).

Japan's enthused individualists

We Were Burning: Japanese Entrepreneurs and the Forging of the Electronic Age

by Bob Johnstone

Basic Books: 1998. 320 pp. \$27.50

Rupert Goodwins

Count the number of LEDs you can see from your desk, or how many liquid-crystal displays you consult each day. Both technologies were invented in the United States; both were commercialized — and how — by Japanese companies. In a world where success in information technology is seen as the only safeguard against economic collapse, the question of how this happened is well worth asking.

It's been asked many times. The accepted wisdom has answered that a coldly efficient Japanese establishment, represented by James Bond villain MITI (the Ministry of International Trade and Industry), has coordinated a decades-long mission to pillage the West's ideas and sell them back at huge profit.

In *We Were Burning*, freelance journalist Bob Johnstone wastes little time in demolishing this xenophobic dogma. As he points out, MITI started out by preventing Sony from making the world's first transistor radio because the company was impertinent enough to act on its own initiative. By the time the ministry's pique had worn off, Texas Instruments had got there first.

Johnstone's thesis throughout this book



Steady nerves: faith in AT&T's discarded 'Picturephone' technology resulted in Sony's camcorder.

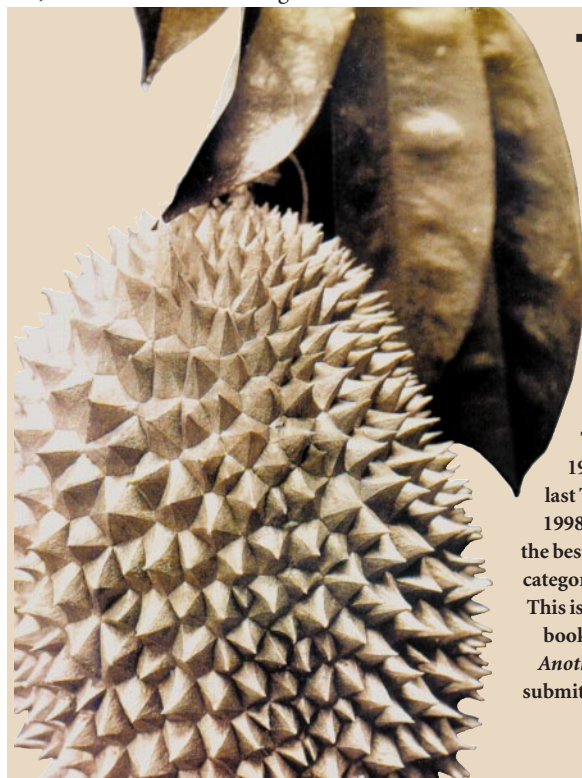
is that enthused individualism, not dispassionate bureaucracy, is the key. Moreover, he demonstrates that clarity and breadth of vision are what drive Japanese success stories: time and again Western inventions are fumbled, then discarded by their inventors, only to be reborn abroad through hard work and commitment. In both cultures the establishment is at best wise enough to leave well alone, at worst an active force for inertia and distraction. In both cultures, exceptional mavericks must set themselves impossible tasks to win out and change our world.

Take the charge-coupled device, the solid-state sensor at the heart of modern

video cameras. Invented by AT&T, it was seen by the company as worthy of development as a key part of the 'Picturephone' initiative. Oddly, people don't want video telephones: AT&T gave up. Iwama Kasuo of Sony, one of the key people behind the company's decision to understand and make transistors, happened upon the technology and saw it as a way to open a whole new market in consumer digital imaging. It took ten years, hundreds of millions of dollars and the very steadiest of nerves: but as a result, Sony created the camcorder market as we know it today. Johnstone tells this story with verve, not stinting on the technicalities but not losing himself or the reader in them either.

For anyone intrigued by or involved with consumer electronics, this book is a treasure-house. It reveals the birth pangs of so many of the things we take for granted — calculators, watches, CD players — but more importantly it describes the heroic (and sometimes tragic) humans who made all this happen. Some of the stories in *We Were Burning* are strong enough to approach mythology, not because they lack rigour or research — quite the opposite — but through their depiction of powerful, intriguing, flawed personalities betting companies and lifetimes on risky, ground-breaking strategies.

A typical tale is that of Yamaha and its 1950s leader Kawakami Gen'ichi: when the company was making 15,000 pianos, he built kilns capable of cooking the wood to make 50,000 instruments. Critics said it would ruin the company. Within ten years, the company was making 100,000 pianos. Kawakami decided to make the best electronic musical instruments, and found that none of his suppliers would make the chips he needed. So he found an abandoned American idea, created his own chip fabrication plant and expected his existing staff — by no means experts, but



The pick of the bunch

This photograph of the distinctively odoured durian fruit, which was produced in the studio of W. L. H. Skeen and Co. between 1862 and 1903, comes from a pioneering collection of photographs of science subjects which makes up *Beauty of Another Order: Photography in Science* by Ann Thomas (Yale University Press, 1997; \$50, £40). It was announced last Tuesday that the book has won the 1998 Kraszna-Krausz Book Award for the best book on photography in the category of craft technology and science. This is an international prize open to books of any language, and *Beauty of Another Order* won out of 270 books submitted from 18 countries.

it's hard to hire in Japan — to learn enough to make it all work. They did, and while US innovator Moog was falling apart, Yamaha ended up with the DX-7, the world's most spectacularly successful synthesizer.

Or take the mad rush in the 1950s to understand and build transistors. One of the tricks is to pull a crystal slowly out of molten metal, surely a job for precisely controlled motors and expensive gearboxes. Yet Japanese researchers at the telephone company NTT roped the crystal to a block of wood which they floated in a bucket of water. Then they drilled a hole in the bottom and waited. There's your crystal; hardly the work of engineering martinets despite the Western image of Japanese conformity. Again and again, Johnstone follows such magnificent stories from unlikely start to the point where they touch our lives. He does this with style and energy, and a richness of detail that's most rewarding.

Yet the book has an air of incompleteness. These are rattlingly good tales, but they remain just that. It's left to the reader to decide how to weave each story into the big picture, which parts are real clues to understanding the Japanese mode of success and which just rumbustious red herrings. One aspect Johnstone fails to cover is what exactly produces consistent results from such a variety of situations. I suspect that the secret lies in the dull old world of making things, something that Johnstone touches on repeatedly but never really follows through.

In the 1980s, I spent some time working in Sinclair Research, a British company that, more than any other in the past 20 years, approached the idiosyncrasy and eccentric determination of the Sonys and Canons. Sinclair developed an unmatched range of new technologies in-house — display devices, semiconductors, software, analog and radio techniques; even (lest we forget) electric vehicles — and fostered an incredibly exciting society of mind as part of the corporate philosophy. *We Were Burning* consistently reminded me of projects and personalities in Sinclair. Yet the company had only one really big money-spinning success — the ZX81 home computer — with even the high-profile products tending to lose money through poor manufacturing or design.

Every company mentioned in the book has had its big failures too. But the ability to actually make the things they design has always been a key aspect of Japanese corporate thinking, even when the investment to make it so is immense. Parenthetically, Johnstone is too harsh on Yamaha when he cites the failed personal computer line as being an example of things going wrong when everything, including the operating system, is developed in-house. Yamaha, like 13 other Japanese computer companies in the early '80s, was persuaded to support the MSX standard developed by an up-and-coming

company called Microsoft. MSX fell flat on its face, taking with it a huge chunk of the nascent Japanese computer industry.

This is a grand book on many levels. It provides a much-needed human dimension to the strange and wonderful threads of technology with which we weave our world, and demonstrates again and again that new ideas enthusiastically promoted will win over protectionism and profiteering. It's not the whole story, but it's intelligent, passionate and very rewarding. Johnstone is to be congratulated on a terrific and timely work. □

Rupert Goodwins is at IT Week,
1 St Kathrine's Way, London E1 9UW, UK

A bestiary of chaos and biodiversity

Biological Exuberance: Animal Homosexuality and Natural Diversity

by Bruce Bagemihl

St Martin's Press: 1998. 735 pp. \$40

Paul H. Harvey

Homosexual and transgender behaviour in the animal kingdom has, over the past 200 years, been frequently dismissed as aberrant. Until now, nobody has attempted to bring together the voluminous scientific literature on the topic. In *Biological Exuberance* Bruce Bagemihl has done an extraordinary job in compiling a vast bestiary. The species-by-species accounts of adult mammals and birds of the same sex courting and mounting each other, living in pairs, defending joint territories and raising young together are fully documented and referenced, and this book should surely become the standard reference work for research on the topics covered.

Bagemihl's 400-page bestiary is preceded by the history of records, studies and evolu-



Two of a kind: it is no rarity to see adult mammals of the same sex courting, mounting each other ...

tionary analyses of homosexual and transgender behaviour. Had it stopped there, I should have applauded a scientific treatise that needed to be written and which, with a few important exceptions, had been accomplished with understanding and rigour. Unfortunately, because more traditional explanations appear insufficient to explain both the high frequencies of these behaviours within species and their recurrence across species, Bagemihl claims to have produced a revolutionary new explanation. In fact he produces no such thing, but instead he rolls together an extraordinary amalgam of concepts ranging in their relevance from the incomprehensible through the unspecified to the inexplicable.

Some standard evolutionary explanations for homosexuality are misunderstood or misrepresented in this book. Oddly for



... grooming each other, living in pairs and raising young together.

someone trained as a biologist and immersed in the topic of evolutionary explanations, the author seems to have missed out on the group-selection debate of the 1960s. The outcome of this was the general acceptance by biologists that evolution by natural selection results in characters that maximize individual reproductive success (or rather, genes that maximize their own reproductive success), even if that is to the detriment of the population.

By contrast, for example, in his survey of evolutionary scenarios, Bagemihl considers seriously the idea that "homosexuality, because it is non-reproductive, acts as a self-regulating mechanism to control a species' population growth". A standard population-genetic analysis would refute the concept that homosexuality evolved to regulate populations, but Bagemihl seems unaware of that. This problem aside, I found the (albeit qualitative) pitching of data against theory to be fair, but analytically naive.

The frequency of homosexuality and transgender behaviour has been underestimated in nature through both bias and ignorance. For example, among many species it is difficult to distinguish between the sexes in the field, so the one on top during mounting is assumed to be the male. The two other possibilities, that the pair is homosexual or that the one on top is the female of a heterosexual pair, have been frequently disregarded.

In many of the best studies, homosexual pairing is observed at a high frequency, in situations where standard evolutionary explanations do not appear to hold. But, as Bagemihl readily admits, the available data are usually insufficient for testing such ideas because they were not collected for that purpose. What he fails to explain, or even mention, is just what has been achieved in the functional analysis of behaviour using, for example, dynamic programming and evolutionarily stable strategy approaches. It is such model-based, analytical methods that should surely form the basis for examining the strategic or evolutionary basis of these behaviours. Speaking personally, it is the general power of evolutionary explanations that has impressed me over the years. If some behaviours are not evolutionarily fine-tuned, this should not be a surprise. However, for Bagemihl, the apparent inadequacy of standard evolutionary theory means that it is time for a revolution.

From where should the revolutionary theory to explain these behaviours come? We should start, we are told, by consulting indigenous knowledge and myths. Well, yes, I agree in part — indigenous peoples have often observed their prey closely for thousands of years, and may have things to tell us about particular species. However, to a greater extent, scientific observations on the peculiarities of the animals involved may help explain the origin of myths. Homosexu-

ality and transgender behaviour are central to many myths, but that does not mean that those myths can explain the reasons for the observed behaviour on which they were based.

After indigenous peoples, using a link I fail to grasp, we are instructed to proceed to new philosophical perspectives in science. The first is post-Darwinian evolution, in which heretical ideas are supposedly being proposed by such luminaries as Mae-Wan Ho and Peter Saunders and becoming generally accepted by evolutionary biologists. That was news to me, and again I could not understand what, if anything, was being claimed.

After accepting post-Darwinian evolution, we should acknowledge its relation to chaos theory, and together they may well explain the diversity of plumage elaboration found in birds. The task is to attempt a chaos-theory analysis of sexual behaviour, whatever that may mean. At this point I was surprised that Gaia had been omitted from the equation — but then she appeared.

Apparently, Gaia "has prompted a rethinking of some of the most basic principles of evolution. Cooperation, in addition to competition, is seen as an important force of evolutionary change" and teaches us that "it may be beneficial for a species or an ecosystem as a whole if some of its members do not procreate".

The logic had lost me — which was, I suppose, the time to introduce the philosophy of Georges Bataille. Apparently, the problem faced by life is the superabundance of energy coming from the Sun. "According to this view, life should in fact be full of 'wasteful', 'extravagant' and 'excessive' activities", and that is why we have homosexuality and transgender behaviour: they are part of biological exuberance. "Contrary to what we have been taught in high school, reproduction is not the ultimate purpose or inevitable

outcome of biology. It is simply one consequence of a much larger pattern of energy 'expenditure' in which the overriding force is the need to use up excess ... Our final resting spot — the concept of Biological Exuberance — lies somewhere along the trajectory defined by these three points (chaos, biodiversity, evolution), although its exact location remains strangely imprecise."

I regret that such a carefully compiled and well-researched volume should fall down so hopelessly when it comes to interpretation. The danger is, of course, that the baby will be thrown out with the bath water. □

Paul H. Harvey is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Worth its salt

Chemins et Savoirs du Sel

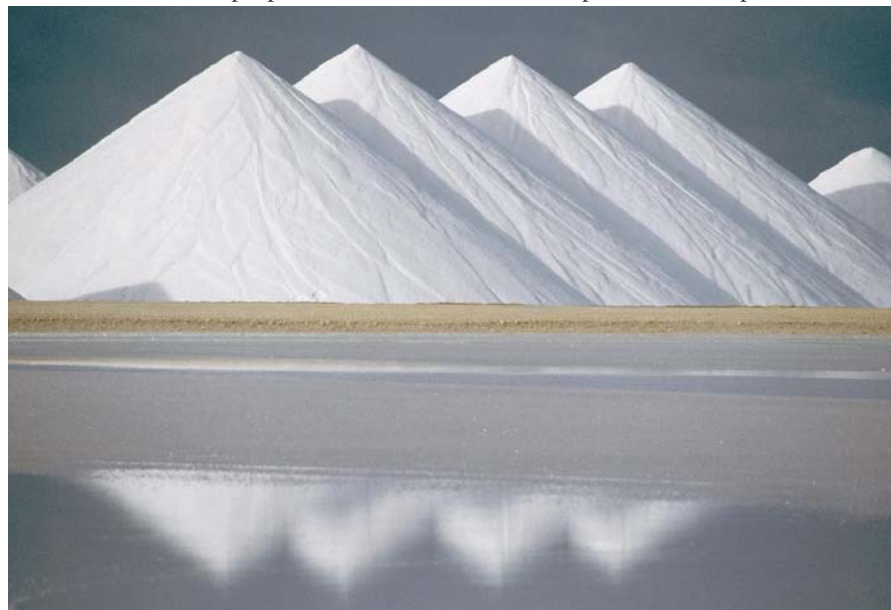
by Pierre Laszlo

Hachette: 1998. 283 pp. FF140

Hervé This

Pierre Laszlo is the author of a wealth of very different books: a very good chemistry textbook; a remarkable small book on alchemy; a larger text showing that chemistry is like a language; a theoretical book about the popularization of science ... and now this volume on salt.

In the introduction to *Chemins et Savoirs du Sel*, Laszlo says that it is an "ignorant treatise", but I prefer to describe it as a walk in a garden where the plants are anecdotes and old wives' tales. There are stories about taxation, episodes in the history of science, reflections on technology and philosophy ... Laszlo points out that Primo Levi did not write chapters about chlorine and sodium in his book *The Periodical Table* (Abacus, £6.99, \$11, pbk), and so he presents us with



Salt of the sea: these sea salt stacks are a monument to the importance of salt throughout history.

information about the two missing chemical elements, mixed together. And as the book examines some stories of oppression — including Auschwitz — it can be said to be related to Levi's famous text.

There are many blooms in this garden. Laszlo tells us how, in 1930, Mahatma Gandhi led the 200-mile Salt March to the sea to collect salt as a symbolic protest against the colonial government's monopoly. He describes the salt production techniques in Guérandes, France. And we learn about the role of salt taxes in Western history.

There are also smaller flowers: the Japanese proverb "*teki ni shio o okuru*", which means to win the victory through fair play, is based on the story of a Japanese general who, during a very long war, sent much-needed salt to his enemy. (The literal meaning of the proverb is "to send salt as a gift to one's enemy".)

The book moves through oppression to pleasure. There is a lot of information here, and Laszlo has extracted the best from many sources. But some aspects of the culinary section are questionable. Laszlo claims that boiling eggs in unsalted water cracks the shell because water enters the egg through osmosis. However, if you compare eggs boiled in salted and unsalted water, both the weight change per egg and the number of cracked eggs are the same. In fact, it is the escaping gases that crack the shell, and all boiled egg lovers should know that salt in boiling water is primarily useful because it diffuses into the egg and makes it tasty.

We are also told that a red wine stain can be avoided if salt is rapidly poured on it, and the theory behind this is described. However, I can refute this from personal experience. At the dinner before the Third International Workshop on Molecular Gastronomy (Ettore Majorana Centre for Scientific Culture, Erice, Sicily, 1995), Pierre-Gilles de Gennes asked me to explain the aim of molecular gastronomy. I said that one of its aims was to check the truth behind culinary maxims.

"Hervé, let us perform a simple experiment," said my old friend Nicholas Kurti (with whom I devised the name molecular gastronomy in 1990). We poured some red wine on a cotton napkin, making four identical stains. After 20 seconds, we poured a large quantity of salt on one of the spots, white wine on the other, water on the third and left the fourth untreated. Two minutes later we removed the salt and compared the stains. All were similar. Since the publication of this book, I have repeated the experiment and found that even if I poured a great excess of salt on a stain, I could not avoid partially staining the napkin.

Another minor error: the book says that the German chemist Justus von Liebig was the inventor of meat extracts, but Antoine-Laurent Lavoisier prepared such tablets in 1783 — and he was not the first, as I found a

recipe for meat extracts in *L'Albert moderne* (Paris, 1770).

I regret that the book is sometimes too intellectual. However, it also makes one think. Here are two examples: Laszlo says that we understand our world if we perceive the questions it is asking us — is the world

asking us something? And "every culturalization of a dish is the result of a violence to nature". Really? □

Hervé This is editor-in-chief of *Pour la Science* (the French edition of *Scientific American*) and is at the *Laboratoire des Interactions Moléculaires*, Collège de France, 75231 Paris, Cedex 05, France.

Science in culture

From destruction to creation

The art of Gustav Metzger

Martin Kemp

Gustav Metzger is infuriating. His programme of 'auto-destructive art', launched in his London manifesto of 1959, leaves behind no collectable artefact, frustrating mercenary dealers and acquisitive museum curators. His 'aesthetic of revulsion', which relishes forms and modes of expression that 'are below the threshold of social acceptance', insults the established criteria of those who aspire to appreciate 'Art'. And his apparently anarchic media, such as the nylon canvases dissolved by acid spray into nothingness, set him up as a prime candidate for ridicule by the popular press and those who decry 'modern art' as fraudulent.

The destructive processes that form one strand of his activity are a response to the events that marked his personal and political life. Born in 1926 as the son of Polish Jews in Nuremberg and exiled with his elder brother to England in 1939, Metzger's student years were ravaged by a war in which his parents were immolated and the art of retaliatory mass destruction was first practised at Hiroshima. In life as in art, he has fermented rebellion. In company with Bertrand Russell and fellow anti-nuclear protesters, he was one of those imprisoned in 1961.

Science — and this is where he qualifies for special mention in *Nature* — has been a constant point of reference in his art. Science is neither automatically demonized nor gratuitously exploited. On one hand, he has savaged the way that scientific technologies have ruthlessly extended man's destructive potential, forcing the old concept of 'nature', ideally inclusive of man, to become the 'environment', which is to be managed by mighty human agencies. And he bears witness, metaphorically in his art, to the awesome 'tearing apart' and 'annihilation' practised by nuclear physics in its dismembering of the atom.

On the other hand, Metzger recognizes that "each disintegration ... leads to the creation of a new form", a transformatory process that leads logically to the other side of his artistic and social coin — 'auto-creative art'. Setting up physico-chemical reactions to produce growth and metamorphosis, auto-creative works exploit the potential of such physical wonders as liquid crystals and advanced technologies, most notably computing, to forge art forms that serve as theatres of natural process. A central ideal is non-predictability, resulting in "a



Gustav Metzger's *Liquid Crystal Environment*, which can be seen at the *Spacex Gallery*, Exeter, UK, from 12 February to 20 March 1999, and at the *Kunsthalle*, Nuremberg, Germany, from 24 June to 12 September 1999.

limitless change of images without image duplication".

His most prominent creations in this auto-creative mode are his liquid-crystal projections, invented in 1965 and first manifested on a huge scale the following year at rock concerts by 'Cream', 'The Who' and 'The Move' at the Roundhouse in London. For his current exhibition, the ponderous 1966 programme of 12 apparatuses operated by a dozen assistants, has been replaced by a computerized system in which six automated projectors, equipped with devices for heating and cooling slides of liquid crystals, cast images through rotating polarized filters onto the walls of the gallery. The endless sequence of infinitely varied configurations and colours pulse with similitudes of organic processes, orchestrated according to different timetables. They simultaneously suggest microbiological metabolisms, geophysical phenomena and the artificial beauty of the kind of abstract art of the 1950s and '60s epitomized by the floating veils of colour on Mark Rothko's canvases.

On one side of his coin, Metzger displays our potential to become the agents of irreversible destruction; on the other he rejoices in our creative integration with the life-giving properties of nature. He is telling us that the choice of on which side the coin falls must not be left to the short-term expediencies of commercial gain. □

Martin Kemp is in the *Department of the History of Art*, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Female development in mammals is regulated by Wnt-4 signalling

Seppo Vainio^{*†}, Minna Heikkilä[†], Andreas Kispert^{*‡}, Norman Chin^{*} & Andrew P. McMahon^{*}

^{*} Department of Molecular and Cellular Biology, The BioLabs, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA

[†] Biocenter Oulu and Department of Biochemistry, Faculties of Science and Medicine, University of Oulu, Linnanmaa, 90570 Oulu, Finland

In the mammalian embryo, both sexes are initially morphologically indistinguishable: specific hormones are required for sex-specific development. Müllerian inhibiting substance and testosterone secreted by the differentiating embryonic testes result in the loss of female (Müllerian) or promotion of male (Wolffian) reproductive duct development, respectively. The signalling molecule Wnt-4 is crucial for female sexual development. At birth, sexual development in males with a mutation in *Wnt-4* appears to be normal; however, *Wnt-4*-mutant females are masculinized—the Müllerian duct is absent while the Wolffian duct continues to develop. *Wnt-4* is initially required in both sexes for formation of the Müllerian duct, then *Wnt-4* in the developing ovary appears to suppress the development of Leydig cells; consequently, *Wnt-4*-mutant females ectopically activate testosterone biosynthesis. *Wnt-4* may also be required for maintenance of the female germ line. Thus, the establishment of sexual dimorphism is under the control of both local and systemic signals.

Early in fetal life, the reproductive system of male and female mammalian embryos consists of an indifferent gonad that is indistinguishable by morphological criteria between the sexes. Adjacent to the gonads are two simple ducts, the Müllerian and Wolffian ducts, the anlagen of female (oviduct, uterus and upper part of the vagina) and male (epididymis, vas deferens and seminal vesicles) reproductive tract, respectively. Thus, the embryo is initially sexually indifferent. Correct sexual development is dependent on differentiation of somatic cell lineages in the gonad as proposed by Burgoyne¹ and the appropriate production, reception and response to gonadal hormones (reviewed in ref. 2).

Male gonadal differentiation is triggered by the action of the Y-chromosome-encoded testis-determining factor (SRY) in the supporting cell lineage of the testis^{1,3–6}. This results in the differentiation of Sertoli cell precursors, and the secretion of a hormone, Müllerian inhibiting substance (MIS), which induces regression of the Müllerian duct^{7–9}. The interstitial cell lineage in the testis then differentiates into Leydig cell precursors, which leads to production of testosterone, thereby promoting development of Wolffian duct derivatives². In the female, absence of MIS permits continued development of the Müllerian duct, whereas absence of testosterone leads to degeneration of the Wolffian duct. Thus, the female pathway of development has traditionally been considered as a default state resulting primarily from the absence of the testis-determining factor and, as a consequence, the failure of MIS and testosterone synthesis. This view is undoubtedly oversimplified. For example, a locus, *Dax-1*, has been implicated in the control of female development^{10–13}.

Here we report that *Wnt-4*, a member of the Wnt family of locally acting cell signals¹⁴, is essential for development of the female, and suppression of the male, reproductive system. At birth, the testes and Wolffian duct derivatives of *Wnt-4*-mutant males appear normal. In contrast, females are masculinized: the Müllerian duct is absent and development of the Wolffian duct resembles that of the male. We propose that *Wnt-4* plays a key role in the female pathway of sexual development by regulating Müllerian duct formation, controlling steroidogenesis in the gonad and possibly supporting oocyte development.

Wnt-4 expression and sexual development

Our earlier studies demonstrated that Wnt-4 signalling is essential

for nephrogenesis in the metanephric kidney¹⁵. In addition to expression of *Wnt-4* in the developing kidney, we observed *Wnt-4* expression in the mesonephros, which participates in gonad formation, in the gonad itself, and in association with the Müllerian duct.

Before gonad formation at 9.5 and 10.5 days post coitum (d.p.c.) in the mouse, *Wnt-4* is expressed along the length of the mesonephros in the mesenchyme, but not in mesonephric tubules (Fig. 1a–c). In the region of the presumptive gonad, expression is also observed in the coelomic epithelium (Fig. 1c). As the gonads emerge at 11.0 d.p.c., *Wnt-4* is expressed in the mesenchyme of the indifferent gonad in both sexes (Fig. 1e, f) and in the mesonephros, but is absent from the mesonephric tubules and Wolffian duct (Fig. 1e, f, and data not shown). A second Wnt family member, *Wnt-6*, is also expressed in the genital ridge (Fig. 1d). At this stage the Müllerian duct is not visible but forms over the next 24 h by an in-folding of *Wnt-4*-expressing coelomic epithelium (data not shown). When sex-specific differentiation of the gonads commences at around 11.5 d.p.c., *Wnt-4* expression is downregulated in the male gonad but is maintained in the female gonad (Fig. 1g–i), and sexually dimorphic expression continues over the period in which the sexes become morphologically distinct (data not shown). As *Wnt-4* is expressed in germ-cell-deficient ovaries of *Steele* (*S1*) mutant embryos, its expression is either partly or completely within somatic cell lineages (Fig. 1j). In contrast to the gonads, *Wnt-4* expression is maintained in the mesonephric mesenchyme of both sexes (Fig. 1g, h) from where some of the gonadal somatic lineages are thought to emerge^{16–18}. In the region of the developing sex ducts, *Wnt-4* is absent from the Wolffian duct epithelium and mesenchyme, but is strongly expressed in mesenchyme cells underlying the newly formed Müllerian duct (Fig. 1k), which itself expresses high levels of a second Wnt family member, *Wnt-7a* (refs 19, 20), and the transcriptional regulator *Pax8* (Fig. 1l). Expression here, and in the ovary, is maintained throughout fetal life (data not shown).

Masculinization of *Wnt-4*-mutant females

The sex-specific regulation of *Wnt-4* in the gonad, and sex-independent expression of *Wnt-4* in the mesenchyme of the mesonephros and Müllerian duct, indicates that *Wnt-4* may be important in development of the reproductive system. To investigate this, we collected the urogenital systems from progeny of heterozygous intercrosses between mice carrying a likely null allele of *Wnt-4* (ref. 15). As mice that are homozygous for this mutation

[‡] Present address: MPI-Institut für Immunobiologie, Abteilung für Molekulare Embryologie, Stübweg 51, 79108 Freiburg, Germany.

die of kidney failure shortly after birth, we examined pups at 18.5 d.p.c. and at birth. Genotype and sex were scored initially by phenotypic inspection of internal reproductive organs and subsequently determined unambiguously by Southern blot analysis with *Wnt-4*- and Y-chromosome-specific probes.

At birth the reproductive organs of male and female pups are quite distinct. The testis is round in appearance, with prominent sex cords, generated by Sertoli cells, ensheathing the germ cells. The proximal Wolffian duct, which corresponds to the developing epididymis, is highly coiled (Fig. 2a, e, i). Testis and ductal development in *Wnt-4*-mutant males are morphologically indistinguishable from wild-type and heterozygous male siblings (Fig. 2b, f, j). Furthermore, expression of the Sertoli cell markers *MIS*²¹ and *Desert hedgehog*²² (*Dhh*), the Leydig cell marker 3β -hydroxysteroid dehydrogenase²³ (3β -HSD), and Wolffian-duct-specific expression of *Sonic hedgehog* (*Shh*)²⁴, are unaltered (data not shown). Thus, male development is unaffected by the loss of *Wnt-4*.

Wild-type females have an elongated ovary which lies closer to the kidney than does the testis (Fig. 2c, g). The ovary is covered by an outer coelomic epithelial capsule (Fig. 2k). In contrast to the

Wolffian duct of the male, the proximal region of the Müllerian duct has only two or three large coils where oviduct formation has started (Fig. 2g). Whereas females that are heterozygous for the *Wnt-4*-mutant allele are similar to their wild-type siblings, the gonad and proximal sex ducts of homozygote females appear masculinized. In most cases the gonad of mutant females is in its correct position (Fig. 2d) but develops closely associated with a fat body that is typical of the male. Furthermore, the gonad is round and unencapsulated as is that of the male (Fig. 2h, l) and the single gonadal-associated duct has a highly convoluted proximal region resembling the developing epididymal region of the male Wolffian duct (Fig. 2d, h, l). In contrast to the internal organs, no masculinization of the external genitalia was evident at birth in *Wnt-4*-mutant females.

To determine the identity of the gonad-associated duct in *Wnt-4*-mutant females, we examined the activity of a *Pax2lacZ* transgene²⁵ which is expressed in the Wolffian duct and its derivatives, and in the collecting duct system and ureter of the kidney. As expected, the transgene is expressed in the Wolffian duct of wild-type and *Wnt-4*-mutant males (Fig. 2m, n) but is absent from wild-type females after

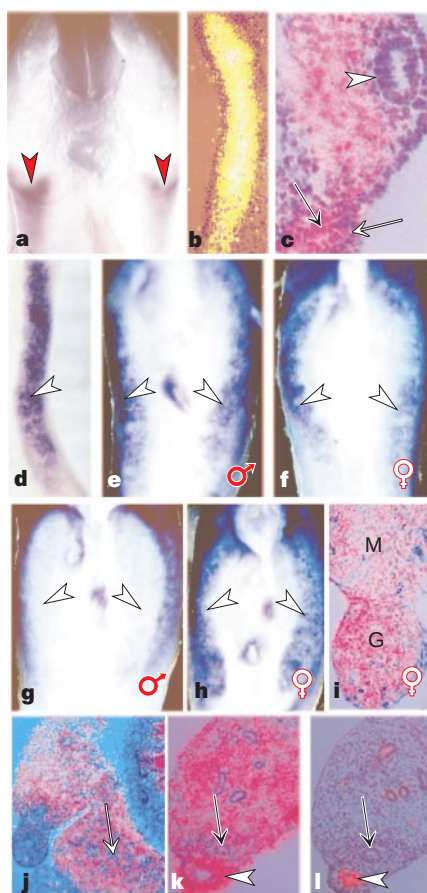


Figure 1 Expression of the *Wnt-4* gene during gonad development. Whole-mount (a, d–h) and section *in situ* hybridization (b, c, i–l) micrographs. At 9.5 d.p.c. (a), *Wnt-4* is present in the mesonephros (red arrowheads) and in the gonadal region at 10 d.p.c. (b, c), both in the coelomic epithelium (white arrow) and mesenchyme (black arrow). At 11.0 d.p.c., both the male (e) and female (f) express *Wnt-4* and a genital ridge marker, *Wnt-6* (d) (arrowheads). At 11.5 d.p.c. *Wnt-4* expression is downregulated in the male (g) and maintained in the female (h, i) (arrowheads). *Wnt-4* is expressed in the ovary of germ-cell-deficient *Steele* mutant (arrow in j). In the sex ducts at 12.5 d.p.c., *Wnt-4* is expressed in the mesenchymal cells surrounding the Müllerian duct (arrowhead in k), whereas the Müllerian duct epithelium expresses *Pax-8* (arrowhead in l). No *Wnt-4* or *Pax-8* expression is observed in association with the Wolffian duct (arrows in c, k, l). M, mesonephros; G, gonad.

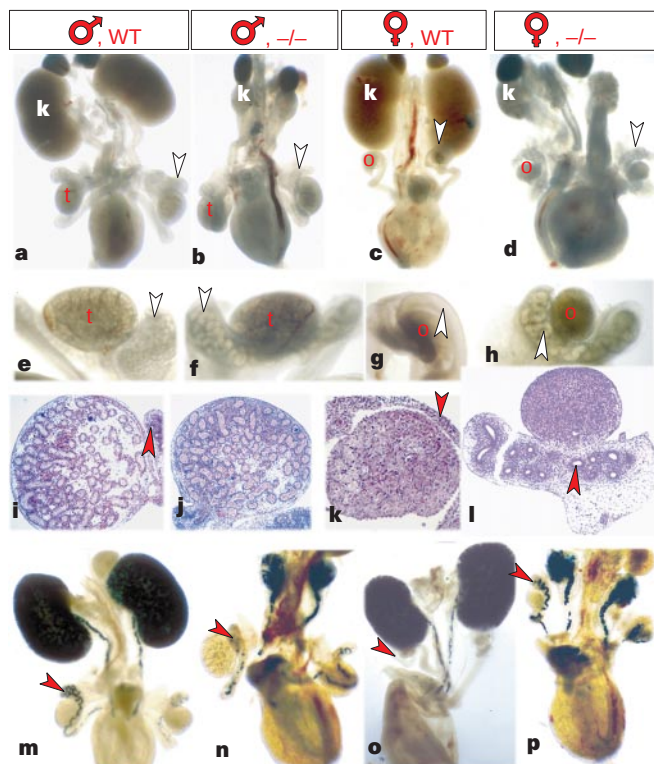


Figure 2 Sex reversal in the reproductive system of *Wnt-4*-mutant females. Urogenital system of newborn mice: a, e, i, m, wild-type males; b, f, j, n, *Wnt-4*-homozygous-mutant males; c, g, k, o, wild-type females; d, h, l, p, *Wnt-4*-homozygous-mutant females. Epididymal region of Wolffian duct in male (a, b, and arrowhead in e, f) and oviduct of Müllerian duct of female (c, and arrowhead in g). In the mutant female, coiling in the sex duct (arrowhead in h) resembles that of the male. Testis develops normally in mutant males (compare i with j). The ovary of mutant females is rounded, lacks a capsule (arrowhead in k) and has a coiled proximal ductal system (arrowhead in l). m–p, *Pax2lacZ* reporter transgene is expressed in the kidney and ureter of all genotypes (m–p) in the Wolffian duct of wild-type (arrowhead, m) and mutant males (arrowhead, n), but also in the single sex duct of the mutant female (arrowhead, p). t, testis; o, ovary; k, kidney.

regression of the Wolffian duct (Fig. 2o). In contrast, the transgene is expressed in the single duct of *Wnt-4*-mutant females (Fig. 2p), indicating that this duct is indeed the Wolffian duct, a conclusion that is further supported by expression of *Shh*, a second Wolffian duct marker (data not shown). Thus, the external appearance of the gonad with its associated fat pad, the absence of a Müllerian duct, and the development of a Wolffian duct are indicative of a reversal of sexual development in *Wnt-4*-mutant females.

Wnt-4 and Müllerian duct formation

Development of the sex ducts in mammals is dependent upon a poorly understood, sex-independent morphogenesis of the ductal systems, and a subsequent hormone-dependent modification that relies on correct differentiation of somatic cell lineages in the gonad^{1,2,16,26}. Failure of Müllerian duct development in the *Wnt-4*-mutant female may therefore reflect either a role for *Wnt-4* in formation of the duct, which would not be sex specific, or a role later in the sex-specific suppression of MIS production in females.

To investigate the fate of the Müllerian duct in *Wnt-4* mutants, we first investigated mutants of both sexes at 14.5 d.p.c., one day before the MIS-induced degeneration of the Müllerian duct in males. Müllerian duct development is identified by the expression of *Wnt-7a*, which is essential for sexually dimorphic development of this duct¹⁹, and *Pax-8*. The Müllerian duct is present in wild-type and heterozygous mice of both sexes as expected (Fig. 3a, c), but is absent from both mutant males and females (Fig. 3b, d). Furthermore, the absence of *Wnt-7a* (Fig. 3e, f) and *Pax-8* (Fig. 3g, h) expression in identifiable Müllerian ducts at 12.5 d.p.c., or at

11.5 d.p.c. when the anterior region of the duct first forms (Fig. 3i–l), indicates that *Wnt-4* signalling is required for the initial stages of ductal morphogenesis. Thus, *Wnt-4* is essential in both sexes for Müllerian duct formation, although it is only in females that its deficiency has a significant consequence.

Steroidogenesis in *Wnt-4*-mutant females

Wolffian duct development depends upon steroid biosynthesis although it is not entirely clear whether testosterone itself, or an intermediate such as androstenedione, represents the early bioactive male steroid²⁷. As *Wnt-4* is not expressed in the Wolffian duct, continued development of this duct in mutant females suggests that *Wnt-4* expression in the ovary may be important in the control of steroidogenesis. Testosterone biosynthesis in Leydig cells of the testes is dependent upon the expression of the several enzymes. For example, 3β -hydroxysteroid dehydrogenase (3β -HSD) converts pregnenolone to progesterone, an intermediate in testosterone biosynthesis. Consequently, a deficiency in 3β -HSD leads to male pseudohermaphroditism²³. 3β -HSD is expressed strongly in Leydig cell precursors in the testes from 12.5 d.p.c. until birth²⁸ (Fig. 4a, d). In the wild-type female, expression is absent from the ovary (Fig. 4b, e), but present in the adrenal gland where 3β -HSD is associated with the synthesis of mineralocorticoids and glucocorticoids²⁸ (data not shown). In contrast, 3β -HSD is expressed in the ovary of *Wnt-4*-mutant females at 14.5 d.p.c. and continues until birth (Fig. 4c, f). A second member of the testosterone biosynthetic pathway, the gene encoding 17α -hydroxylase/C₁₇₋₂₀ lyase (*P450c17*), which catalyses two reactions

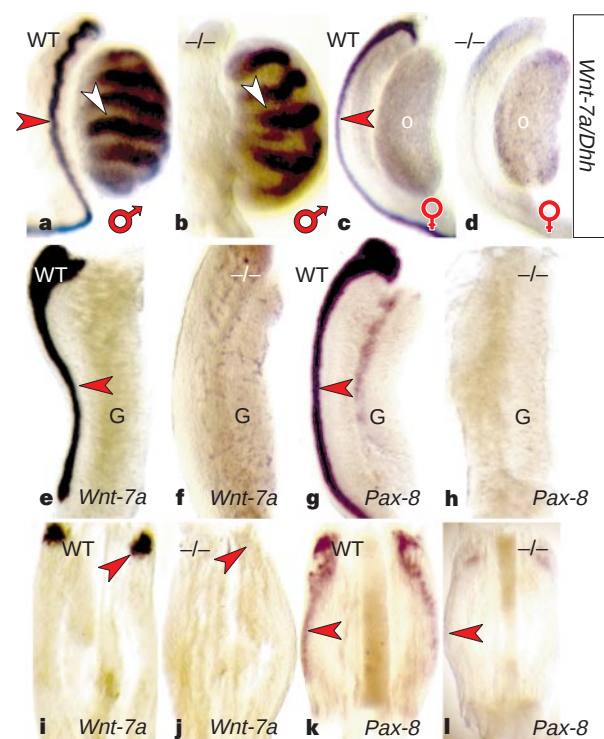


Figure 3 Sex-independent regulation of Müllerian duct formation by *Wnt-4*. Whole-mount *in situ* hybridization micrographs. *Wnt-7a* and *Pax-8*, markers of Müllerian duct development (filled arrowheads in **a**, **c**, **e**, **g**) and *Dhh*, a marker of Sertoli cell precursors within the testes cords (white arrowhead in **a** and **b**). In contrast to the wild-type males (**a**) and females (**c**), neither *Wnt-4*-mutant males (**b**) nor females (**d**) have a Müllerian duct. **a**, **d**, 14.5 d.p.c. Analysis of *Wnt-7a* (**e**, **f**) and *Pax-8* (**g**, **h**) expression at 12.5 d.p.c. demonstrates a Müllerian duct in wild-type (arrowhead in **e**, **g**) but not in *Wnt-4* mutant (**f**, **h**) animals. Activation of *Wnt-7a* (**i**, **j**, arrowhead) and *Pax-8* (**k**, **l**, arrowhead) expression associated with the initial stages of Müllerian duct development is absent in *Wnt-4* mutants at 11.5 d.p.c.

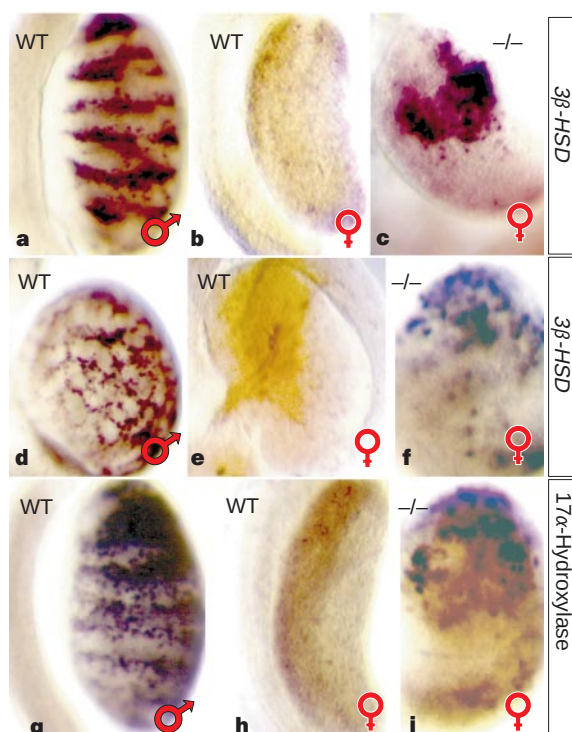


Figure 4 Ectopic expression of Leydig cell markers, 3β -HSD and 17α -hydroxylase genes, in the gonad of *Wnt-4*-mutant females. Whole-mount *in situ* hybridization micrographs. 3β -HSD and 17α -hydroxylase genes are expressed in Leydig cells in the male (**a**, **d**, **g**) but not the female (**b**, **e**, **h**) gonad. Patchy ectopic expression of 3β -HSD (**c**, **f**) and 17α -hydroxylase (**i**) is observed in the gonad of *Wnt-4*-mutant females. **a**–**c**, **g**, **h**, 14 d.p.c.; **d**, **e**, **f**, **i**, neonate.

converting progesterone into androstenedione, shows testis-specific expression in Leydig cell precursors during normal fetal development²⁹ (Fig. 4g, h). However, we also observed patchy ectopic activation of *P450c17* in the gonad of *Wnt-4*-mutant females from 14.5 d.p.c. until birth (Fig. 4i, and data not shown). Thus, in the absence of *Wnt-4* signalling, steroidogenesis is initiated in the ovary. Consequently, it is likely that biosynthesis and secretion of testosterone promotes Wolffian duct development in *Wnt-4*-mutant females, although the levels of hormone appear to be insufficient to masculinize external genitalia. Alternatively, efficient export to these target tissues may require a male-specific vasculature which is not formed in *Wnt-4* mutants.

As steroidogenesis in the fetal gonad is normally associated with initiation of Leydig cell development in the testis, these results indicate that female-specific *Wnt-4* signalling is required to suppress inappropriate differentiation of Leydig cell precursors in the fetal ovary. However, as both genes are also expressed in the mature ovary, participating in the biosynthesis of oestradiol, it is possible that *Wnt-4* may operate differently, preventing premature differentiation of thecal and/or granulosa cells. We attempted to distinguish between these two possibilities by examining expression of the gene encoding a third steroidogenic enzyme, the type-III isoform of 17 β -HSD, which is reported to be exclusively testis specific³⁰. Although, type-III 17 β -HSD is expressed ectopically in the *Wnt-4* mutant ovary at birth, *in situ* hybridization to neonatal testis samples indicates that type-III 17 β -HSD is not expressed in Leydig cells but is most likely a Sertoli cell product (data not shown).

This result indicates that the supporting cell lineage in the ovary of *Wnt-4* mutants may have undergone a primary sex-reversal which may subsequently lead to a secondary change in interstitial cell fates. To investigate this, we examined expression of *Dhh*, *MIS* and *Sox-9* (ref. 31), all of which are pre-Sertoli cell markers that are expressed specifically in the supporting cell lineage of the testes. None of these genes is ectopically activated in the ovaries of *Wnt-4* mutants before normal loss of the Wolffian duct (Fig. 3a–d, and data not shown). This result, together with the absence of morphologically identifiable sex cords when ectopic steroidogenic gene expression is first observed, indicates that the ovaries of *Wnt-4* mutants have not undergone a primary sex reversal leading to

Sertoli cell development. In addition, we continued to detect expression of *Dax-1*, which is implicated in the female pathway of sex determination¹³ (data not shown).

Wnt-4 supports oocyte development

In addition to ectopic steroidogenesis, we also observed a marked reduction in oocyte development in mutant females. Germ cells enter the gonads between 10.5 and 12.5 d.p.c. (reviewed in ref. 32). In the female, germ cells enter meiotic prophase and progress to the diplotene stage by birth. In the male, after an initial proliferative phase they enter mitotic arrest. *In situ* hybridization at 14.5 d.p.c. with *Oct-4*, a germ cell marker, indicated normal numbers of germ cells in both the ovary and testes of *Wnt-4* mutants (data not shown). Consequently, the early phases of germ cell migration and proliferation are unaffected. In contrast, whereas wild-type or heterozygous females have many readily detectable, meiotic-stage cortical oocytes at birth (Fig. 5a, b), the ovaries of *Wnt-4*-mutant females have only a few oocytes and these are in the process of degenerating (Fig. 5c, d). We confirmed the specific loss of oocytes by staining testes and ovaries with an antibody recognizing the germ cell nuclear antigen (GCNA-1)³³. Although no differences are observed in the testes (Fig. 5e, f), *Wnt-4*-mutant ovaries have less than 10% of the oocytes scored in wild-type or heterozygous siblings (Fig. 5g, h). As chronic administration of testosterone does not impair oocyte development³⁴, our results indicate that the loss of oocytes is probably not a secondary consequence of altered interstitial cell fates, but could reflect a more direct role for *Wnt-4* in maintaining development of the female germ line, a possibility that warrants further investigation.

Several lines of evidence indicate that the loss of oocytes leads to a secondary sex-reversal of supporting cell lineages^{34–36}. Thus, continued interactions between the oocyte and supporting cells, the nature of which are unclear, are necessary for maintaining follicle cell and suppressing Sertoli cell fates. In support of this view, we find that a few sex-cord-like structures have emerged in oocyte-depleted ovaries of *Wnt-4* mutants at birth (Fig. 5i). Moreover, in contrast to the wild-type ovary, expression of the Sertoli cell markers *MIS* (Fig. 5j, k) and *Dhh* (Fig. 5l) is detected. Thus although no apparent sex reversal is observed in the supporting cell lineage of the *Wnt-4*-

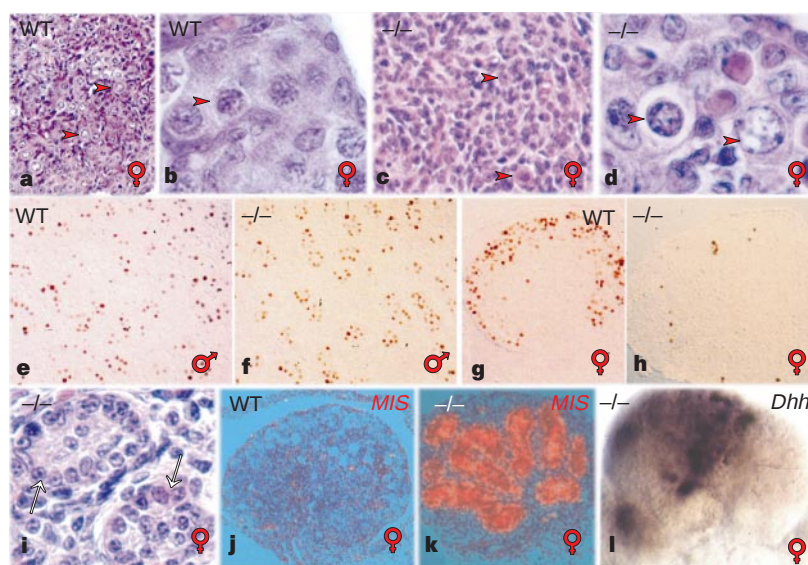


Figure 5 Oocyte development is dependent on *Wnt-4*. Low-power (a, c) and high-power (b, d) views through wild-type (a, b) and *Wnt-4*-mutant (c, d) ovaries at birth reveals a depletion of oocytes (arrowheads) in the ovary of mutants. Some degenerating oocytes are visible in the mutant (arrowheads in c, d). GCNA-1 expression in germ cells in the testis (e, f) and ovary (g, h) of wild-type (e, g) and

Wnt-4 mutants (f, h) reveals loss of oocytes in mutants. Sex cords similar to those in the male can be seen in the ovary of *Wnt-4* mutants (arrows in i). Furthermore, unlike the wild-type ovary (j), the Sertoli cell markers *MIS* (k) and *Dhh* (l) are ectopically expressed in sex cords of *Wnt-4* mutants at birth and indicate a partial somatic sex-reversal in the absence of oocytes.

mutant ovary earlier in development, some supporting cells adopt characteristics of Sertoli cells after the loss of oocytes.

Discussion

Our results have revealed a hitherto unappreciated role for Wnt signalling in female development. Surprisingly, *Wnt-4* appears to be required for three distinct aspects of the female pathway—formation of the Müllerian duct, differentiation of the interstitial cell lineage and oocyte development. The failure to identify the Müllerian duct in either sex between 11.5 and 12.5 d.p.c. indicates that *Wnt-4* regulates morphogenesis of this duct from the coelomic epithelium. It is not clear whether this process bears mechanistic similarities to the action of *Wnt-4* in formation of kidney tubules¹⁵. Once formed, a second *Wnt* member, *Wnt-7a*, regulates sex-specific development of the Müllerian duct and its derivatives^{19,20}, and so there are at least two phases to *Wnt* action in ductal organogenesis.

In the ovary, ectopic activation of genes that encode enzymes in the testosterone biosynthetic pathway, whose gonadal expression is normally restricted to the testis at these stages, indicates that during normal development, ovary-specific expression of *Wnt-4* suppresses Leydig cell development. Furthermore, the results indicate that steroid cell precursors are most likely present in both the male and female gonad and are not entering in response to a sex-specific signal initiated by the Sertoli cells, as appears to be the case with some somatic lineages in the testis. As *Wnt-4* expression is lost in the testis in conjunction with male sex determination, *Wnt-4* may be negatively regulated by the male sex-determining pathway. At present we are unable to determine with the available reagents which component of the somatic cell lineages expresses *Wnt-4*, so it is not clear whether *Wnt-4* could be a direct target of the sex-determining transcriptional regulators that initiate differentiation of the supporting cell lineage.

Outside the gonad, *Wnt-4* is also expressed in the adrenal cortex of both sexes, which is consistent with a broader role for *Wnt-4* in the regulation of steroidogenesis. Two nuclear factors, *SF-1* and *Dax-1*, demonstrate a connection between sex-specific differentiation in the gonad and regulation of adrenal gland development^{11,13,37,38}. Finally, the early loss of the female germ line indicates that *Wnt-4* may be important in the post-meiotic maintenance of oocytes. The requirement for somatic-cell-derived factors in early stages of oocyte development has not been rigorously addressed. However, growth and survival of oocytes in the adrenal gland of both sexes has been reported³⁹. As this would place oocytes in association with *Wnt-4*-expressing cells, it is possible that *Wnt-4* may support oocyte development in this ectopic location. □

Methods

Genotyping and sex typing of embryos. The *Wnt-4*-mutant mouse line and *Wnt-4*-mutant embryos were genotyped as reported earlier¹⁵. For sex typing, embryos were collected separately, DNA was isolated by routine methods⁴⁰, either from the yolk sac or from the remainder of the embryo after removal of the urogenital system, and then digested with *EcoRI*. Y-chromosome-specific sequences were detected by Southern blot hybridization with the Y-chromosome repetitive probe, pY353/B¹. After genotyping and sex typing, samples were pooled accordingly. The *Pax-2* transgenic mouse line has been reported earlier²⁵ and was typed using a *LacZ* complementary DNA probe.

Histology, *in situ* hybridization, β -galactosidase and antibody staining. Histological procedures, whole-mount and section *in situ* hybridization and β -galactosidase staining were done according to routine procedures⁴⁰. Anti-GCNA-1 antibody was used at a concentration of 1:1,000 as described³³ and antibody binding was visualized using a Vectastain ABC kit according to the manufacturer's recommendations.

Received 27 April; accepted 21 December 1998.

1. Burgoyne, P. S. Role of the mammalian Y chromosome in sex determination. *Phil. Trans. R. Soc. Lond. B* **322**, 63–72 (1988).
2. Goodfellow, P. N. & Lovell-Badge, R. SRY and sex determination in mammals. *Annu. Rev. Genet.* **27**, 71–92 (1993).

3. Sinclair, A. H. *et al.* A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **366**, 240–244 (1990).
4. Gubbay, J. *et al.* A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**, 245–250 (1990).
5. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. Male development of chromosomally female mice transgenic for *SRY*. *Nature* **351**, 117–121 (1991).
6. Palmer, S. J. & Burgoyne, P. S. In situ analysis of fetal, prepubertal and adult XX–XY chimaeric mouse testes: Sertoli cells are predominantly, but not exclusively, XY. *Development* **112**, 265–268 (1991).
7. Jost, A. Problems of fetal endocrinology: the gonadal and hypophyseal hormones. *Recent Prog. Horm. Res.* **8**, 379–418 (1953).
8. Cate, R. L. *et al.* Isolation of the bovine and human genes for Müllerian-inhibiting substance and expression of the human gene in animal cells. *Cell* **45**, 685–698 (1986).
9. Behringer, R. R., Finegold, M. J. & Cate, R. L. Müllerian-inhibiting substance function during mammalian sexual development. *Cell* **79**, 415–425 (1994).
10. Bardoni, B. *et al.* A dosage sensitive locus at chromosome Xp21 is involved in male to female sex reversal. *Nature Genet.* **7**, 497–501 (1994).
11. Zanaria, E. *et al.* An unusual member of the nuclear hormone receptor superfamily responsible for X-linked adrenal hypoplasia congenita. *Nature* **372**, 635–641 (1994).
12. Muscatelli, F. *et al.* Mutations in the *DAX-1* gene gives rise to both X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism. *Nature* **372**, 672–676 (1994).
13. Swain, A., Narvaez, V., Burgoyne, P., Camerino, G. & Lovell-Badge, R. *Dax1* antagonizes *SRY* action in mammalian sex determination. *Nature* **391**, 761–767 (1998).
14. Cadigan, K. M. & Nusse, R. Wnt signaling: a common theme in animal development. *Genes Dev.* **11**, 3286–3305 (1997).
15. Stark, K., Vainio, S., Vassileva, G. & McMahon, A. P. Epithelial transformation of metanephric mesenchyme in the developing kidney regulated by *Wnt-4*. *Nature* **372**, 679–683 (1994).
16. McLaren, A. Development of the mammalian gonad: the fate of the supporting cell lineage. *BioEssays* **13**, 151–156 (1991).
17. Buehr, M., Gu, S. & McLaren, A. Mesonephric contribution to testis differentiation in the fetal mouse. *Development* **117**, 273–281 (1993).
18. McLaren, A. Gonad development: assembling the mammalian testis. *Curr. Biol.* **8**, 175–177 (1998).
19. Parr, B. A. & McMahon, A. P. Sexually dimorphic development of the mammalian reproductive tract requires *Wnt-7a*. *Nature* **395**, 707–710 (1998).
20. Miller, C. & Sassoon, D. A. *Wnt-7a* maintains appropriate uterine patterning during the development of the mouse female reproductive tract. *Development* **125**, 3201–3211 (1998).
21. Münsterberg, A. & Lovell-Badge, R. Expression of the mouse anti-Müllerian hormone gene suggests a role in both male and female sexual differentiation. *Development* **113**, 613–624 (1991).
22. Bitgood, M. J., Shen, L. & McMahon, A. P. Sertoli cell signaling by Desert hedgehog regulates the male germ line. *Curr. Biol.* **6**, 298–304 (1996).
23. Morel, Y. *et al.* Structure-function relationships of β -hydroxysteroid dehydrogenase: contribution made by the molecular genetics of β -hydroxysteroid dehydrogenase deficiency. *Steroids* **62**, 176–184 (1997).
24. Bitgood, M. J. & McMahon, A. P. *Hedgehog* and *Bmp* genes are co-expressed at many diverse sites of cell-cell interaction in the mouse embryo. *Dev. Biol.* **172**, 126–138 (1995).
25. Kispert, A., Vainio, S., Shen, L., Rowitch, D. H. & McMahon, A. P. Proteoglycans are required for maintenance of *Wnt-11* expression in the ureter tips. *Development* **122**, 3627–3637 (1996).
26. Nordqvist, K. & Lovell-Badge, R. Setbacks on the road to sexual fulfillment. *Nature Genet.* **7**, 7–9 (1994).
27. Andersson, S. *et al.* Molecular genetics and pathophysiology of 17 β -hydroxysteroid dehydrogenase 3 deficiency. *J. Clin. Endocrinol. Metab.* **81**, 130–136 (1996).
28. Nordqvist, K. & Tohonen, V. A mRNA differential display strategy for cloning genes expressed during mouse gonad development. *Int. J. Dev. Biol.* **41**, 627–638 (1997).
29. Greco, T. L. & Payne, A. H. Ontogeny of expression of the genes for steroidogenic enzymes P450 side-chain cleavage, β -hydroxysteroid dehydrogenase, P450 17 α -hydroxylase/C₁₇₋₂₀ lyase and P450 aromatase in fetal mouse gonads. *Endocrinology* **135**, 262–268 (1994).
30. Mustonen, M. V., Poutanen, M. H., Isomaa, V. V., Viikko, P. T. & Viikko, R. K. Cloning of mouse 17 β -hydroxysteroid dehydrogenase type 2, and analyzing expression of the mRNAs for types 1, 2, 3, 4 and 5 in mouse embryos and adult tissues. *Biochem. J.* **325**, 199–205 (1997).
31. Kent, J., Wheatley, S. C., Andrews, J. E., Sinclair, A. H. & Koopman, P. A male-specific role for *SOX9* in vertebrate sex determination. *Development* **122**, 2813–2822 (1996).
32. McLaren, A. in *Organization of the Early Vertebrate Embryo* (eds Zagris, N., Duprat, A. M. & Durston, A. J.) 1–9 (Plenum, New York, 1995).
33. Enders, G. C. & May, J. J. II Developmentally regulated expression of a mouse germ cell nuclear antigen examined from embryonic day 11 to adult in male and female mice. *Dev. Biol.* **163**, 331–340 (1994).
34. Lyet, L. *et al.* Ontogeny of reproductive abnormalities induced by deregulation of anti-Müllerian hormone expression in transgenic mice. *Biol. Reprod.* **52**, 444–454 (1995).
35. Vigier, B., Watrin, F., Magre, S., Tran, D. & Josso, N. Purified bovine AMH induces a characteristic freemartin effect in fetal rat prospective ovaries exposed to it in vitro. *Development* **100**, 43–55 (1987).
36. Behringer, R. R., Cate, R. L., Froelick, G. J., Palmiter, R. D. & Brinster, R. L. Abnormal sexual development in transgenic mice chronically expressing Müllerian inhibiting substance. *Nature* **345**, 167–170 (1990).
37. Luo, X., Ikeda, Y. & Parker, K. L. A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell* **77**, 481–490 (1994).
38. Sadovsky, Y. *et al.* Mice deficient in the orphan receptor steroidogenic factor 1 lack adrenal glands and gonads but express p450 side-chain-cleavage enzyme in the placenta and have normal embryonic serum levels of glucocorticoids. *Proc. Natl Acad. Sci. USA* **92**, 10939–10943 (1995).
39. Zamboni, L. & Upadhyay, S. Germ cell differentiation in mouse adrenal glands. *J. Exp. Zool.* **228**, 178–193 (1983).
40. Hogan, B., Beddington, R., Constantini, F. & Lacy, E. *Manipulating the Mouse Embryo: A Laboratory Manual* 2nd edn. (Cold Spring Harbor Press, New York, 1994).
41. Bishop, C. E., Boursot, P., Baron, B., Bonhomme, F. & Hatat, D. Most classical *Mus musculus domesticus* laboratory mouse strains carry a *Mus musculus musculus* Y chromosome. *Nature* **315**, 70–72 (1985).

Acknowledgements. We thank P. Donahue, R. Lovell-Badge, R. Viikko, H. Peltoketo, P. O'Shaughnessy, R. Maas, J. Smith, P. Gruss and G. Enders for supplying cDNAs used in hybridization analysis; G. Enders for supplying the GCNA-1 antibody; D. Rowitch for help in generating the *Pax-2* transgenic mouse line; B. Klumpp for technical assistance; and A. McLaren, and R. Lovell-Badge for discussions. S.V. was supported by the academy of Finland, Finnish Cultural and Jusélius Foundations, and European Molecular Biology Organization; A.K. was the recipient of a long-term fellowship from Human Frontier Science Program Organization and Otto-Hahn Fellowship of the Max-Planck Society; and work in A.P.M.'s laboratory was funded by a grant from the NIDDK at the NIH.

Correspondence and requests for materials should be addressed to A.P.M. (e-mail: amcmahon@biosun.harvard.edu).

Coherence and single-particle excitations in the high-temperature superconductors

Guy Deutscher

School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel

The 'pseudogap' observed in the electron excitation spectrum of underdoped copper oxide superconductors has become the focus of considerable attention in the field of high-temperature superconductivity. In conventional superconductors, described by 'BCS' theory, an energy gap appears at the superconducting transition temperature (T_c); the pseudogap, in contrast, is observed well above T_c (ref. 1) and can be large compared to the conventional BCS gap²⁻⁴. Here I compare gap energies, measured by different experimental techniques, for the copper oxide superconductors and show that these reveal the existence of two distinct energy scales: Δ_p and Δ_c . The first, determined either by angle-resolved photoemission spectroscopy or by tunnelling, is the single-particle excitation energy—the energy (per particle) required to split the paired charge-carriers that are required for superconductivity. The second energy scale is determined by Andreev reflection experiments, and I associate it with the coherence energy range of the superconducting state—the macroscopic quantum condensate of the paired charges. I find that, in the overdoped regime, Δ_p and Δ_c converge to approximately the same value, as would be the case for a BCS superconductor where pairs form and condense simultaneously. But in the underdoped regime, where the pseudogap is observed, the two values diverge and Δ_p is larger than Δ_c . Models that may provide a framework for understanding these results involve the existence of pairing above the condensation temperature, as might occur in a crossover from BCS to Bose–Einstein condensation behaviour⁵ or from the formation of striped phases⁶.

Fluctuations of the superconducting order parameter are negligible in the low-temperature superconductors, but have been predicted to be large in the high- T_c copper oxides, due either to their short coherence length⁷ or to their large penetration depth⁶. The short coherence length may signal a crossover from BCS to Bose–Einstein condensation⁵, in which local pairs can form above the condensation temperature: local pair formation is driven by a gain in potential energy. Alternatively, the large penetration depth may be associated with a phase separation between metallic regions and insulating ones having a spin gap⁶. This separation reduces the kinetic energy, and is accompanied by the formation of a pairing amplitude through a proximity effect between the metallic and insulating stripes, but with large phase fluctuations. In both cases, a pairing amplitude will appear above T_c . These models may explain the formation of the pseudogap in the underdoped regime, although alternative explanations have also been proposed, such as the opening up in the density of states of a gap of magnetic origin⁸.

The relationship between T_c and the penetration depth λ (and hence the superfluid density), shown by Uemura⁹, does indicate the existence of strong phase-fluctuation effects. It has been interpreted as being due to a weak phase stiffness⁶, or to a BCS to Bose–Einstein crossover⁹, the latter also being suggested by an isotope effect on λ (ref. 10). However, no distinctive spectroscopic features of the condensed state, which would be characteristic of any of the above-mentioned mechanisms of pseudogap formation, have been proposed up to now.

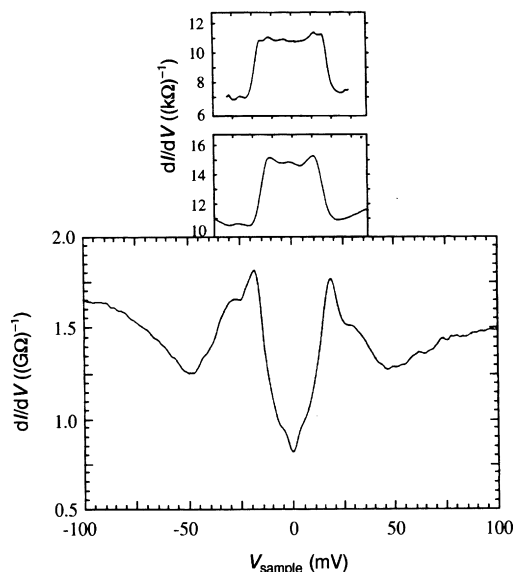


Figure 1 Andreev and tunnelling spectroscopies in optimally doped YBCO. Upper panels, gap edge (~ 20 meV), as seen from Andreev enhanced sub-gap conductance, from an Au-optimally doped textured YBCO point contact orientated along one of the in-plane orthorhombic principal axis (from ref. 12). Lower panels, STM tunnelling characteristic from an optimally doped YBCO single crystal¹¹. The position of the main tunnelling conductance peak matches the Andreev gap edge, in accordance with the predictions of BCS theory.

I propose here a direct test of the nature of the condensation, which consists of comparing two kinds of determination of the gap: one that is sensitive to the single-particle excitation energy Δ_p , and the other to the energy coherence range Δ_c . In a BCS condensate, these two energy scales are identical, basically because the pairs form and condense simultaneously at T_c , owing to the large coherence length. If there are strong phase fluctuations, the single-particle excitation energy has no relation to the condensation temperature, and the two scales may be different. They will evidently also be different if Δ_p has no relation to a pairing amplitude. As I now show, experiments indicate that the two energy scales converge in the overdoped regime, but depart from each other in the underdoped regime, Δ_p becoming larger and Δ_c smaller. It immediately follows that, in the underdoped case, the superfluid condensation is not in the BCS limit.

Single-particle (Giaever) tunnelling and photoemission experiments are two methods that can be used to determine Δ_p . In a system of two-fermion composite bosons, for instance, such experiments measure the energy $2\Delta_p$ necessary to break the paired fermions into single particles (Δ_p per particle), irrespective of whether they are, or are not, in a condensed state. On the other hand, the Josephson effect only takes place in the condensed state, because it is a manifestation of macroscopic quantum coherence (and not just of the existence of a pairing amplitude). The Josephson effect will disappear above the condensation temperature, while the single-particle energy gap can persist to those higher temperatures if the pairs are nevertheless present. Likewise, I expect that Andreev reflections—in which an incoming electron from the normal side of a normal/superconducting contact is reflected as a hole along the same trajectory while a Cooper pair flows on the superconducting side—will only lead to enhanced conductance of the contact (ideally by a factor of 2) if the pairs are in a condensed state. In the case of strong phase fluctuations, one may expect that these two experiments (Josephson and Andreev) will give an energy scale Δ_c (within numerical factors of the order of unity; the product of the critical current of the junction I_c and its normal state resistance R_n for the Josephson effect, and the bias value up to which Andreev

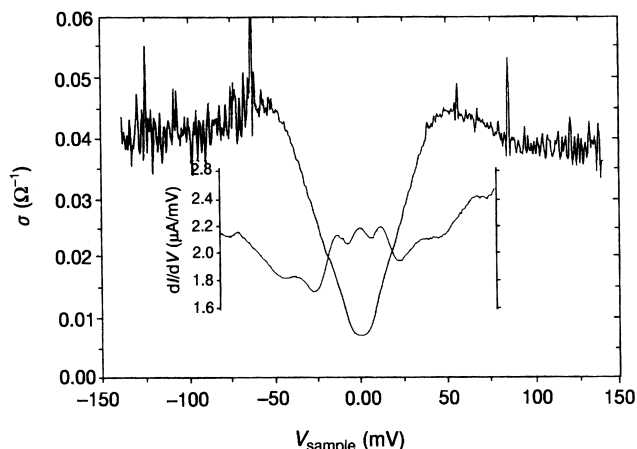


Figure 2 Andreev and tunnelling spectroscopies in underdoped YBCO. The Andreev gap edge at ~ 13 meV (ref. 14), shown in the inset, is now much smaller than the bias at the tunnelling conductance (σ) maxima, at about 50 meV (from ref. 7). Note the weak structure seen in the tunnelling conductance at the Andreev gap edge.

reflections are observed) that is different from (and smaller than) Δ_p . In a BCS condensate, on the other hand, these two energy scales are identical, as already mentioned. I am not aware of any specific calculations of Josephson or Andreev effects, in the presence of strong phase fluctuations, that go beyond these general remarks.

I have compared Giaever (tunnelling) and Andreev gap determinations on YBCO, LSCO and BSCCO, at various doping levels, and have found a systematic difference between these determinations in the underdoped regime (YBCO is $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$, LSCO is $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ and BSCCO is $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$). (The advantage of Andreev over Josephson experiments to check the coherence energy range is that the former do not suffer from the many extrinsic effects that can reduce the Josephson critical current.)

As an example, Fig. 1 shows the results of both tunnelling and Andreev experiments on optimally doped YBCO. The tunnelling data were obtained by scanning tunnelling microscopy (STM) on a single crystal along the c -axis direction¹¹, for which the conductance (dI/dV) measures the average density of states. The position of the main conductance peak then gives the gap value if, as is now generally believed, it has a d -wave symmetry. The Andreev data were obtained by point contact spectroscopy on a melt textured sample, along the a (or b) axis direction¹². For these orientations, the return to the normal-state conductance occurs at the gap edge, whether it has s -wave or d -wave symmetry¹³. The position of the main peak in the STM conductance is in good agreement with that of the sharp Andreev gap edge, both being close to 20 meV. From this agreement, I conclude that YBCO at optimum doping is in the BCS condensation regime. This observation fits with the STM observation of bound states in vortices¹¹, another manifestation of the number of pairs per coherence volume being significantly larger than unity.

By contrast, I show in Fig. 2 tunnelling and Andreev data on underdoped a -axis-orientated YBCO films. The position of the tunnelling conductance peak is now at a much larger bias (50 meV; ref. 3) than that of the Andreev gap edge (13 meV; ref. 14). Comparing the underdoped to the optimally doped data, Δ_p has increased while Δ_c has decreased. The ratio between the two scales is now of the order of 4. In underdoped BSCCO, Δ_p is also close to

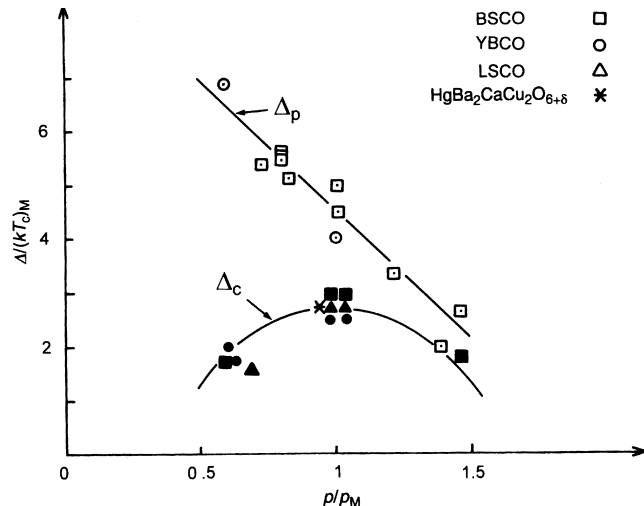


Figure 3 The single-particle excitation energy gap, Δ_p , and the coherence energy range, Δ_c , for a number of copper oxides. Results are shown as a function of the normalized doping level (p/p_M) and gap values are normalized by the respective values of kT_c . Δ_p values are from tunnelling and photoemission data; Δ_c values are from Andreev, penetration depth and Raman B_{2g} data. When not available, p has been calculated from the relation $(T/T_{cMAX}) = 1 - 82.6(p - 0.16)^2$. Here p_M is the doping level at the maximum value of T_c , T_{cMAX} . Open symbols, photoemission and tunnelling data; filled symbols, Andreev, penetration depth and Raman data. For detailed references, see Table 1.

50 meV (ref. 15) and Δ_c to 13 meV (ref. 16). (I consider that the latter result was obtained on an underdoped junction, because it had a T_c of only 70 K according to Fig. 5 in ref. 16. Such underdoping is known to occur in junctions prepared through contact with easily oxidized metals, such as Pb; ref. 3). In underdoped LSCO, a tunnelling gap close to 15 meV has been reported^{17,18}, compared to an Andreev scale of ~ 3 meV (ref. 19). Therefore, in the underdoped case, $\Delta_p > \Delta_c$ and condensation is not in the BCS limit.

Penetration-depth measurements at low temperature have given a slope ($d\lambda/dT$) that increases in the underdoped regime, being inversely proportional to T_c (refs 20, 21). This behaviour is inconsistent with a BCS interpretation of the pseudogap becoming a d -wave BCS gap below T_c : as the pseudogap increases in the underdoped regime, the slope should decrease, according to d -wave theory which predicts that the slope should be inversely proportional to the gap²¹. But it is consistent with the values of Δ_c obtained from Andreev experiments, the energy scale following T_c in both cases. It has also been remarked that the position of the peak in Raman scattering in the B_{2g} polarization scales with T_c in the underdoped regime²². What the penetration depth and Raman B_{2g} have in common is that they are insensitive to the gap structure near the $(\pi, 0)$ and equivalent points on the Fermi surface. It is known from photoemission experiments that pseudogap formation, apparently related to the anomalous concentration dependence of Δ_p , is localized near the $(\pi, 0)$ points².

To emphasize the difference in behaviour between Δ_p and Δ_c as a function of doping, data from the various kinds of measurements cited here are shown Fig. 3. In Table 1, values of the strong coupling ratio ($2\Delta/kT_c$) are given for the data shown in Fig. 3. It varies from 6 to 20 for $\Delta = \Delta_p$, and between 4 and 6 for $\Delta = \Delta_c$. One may wonder whether Δ_c is not simply the 'true' superconducting gap, with Δ_p bearing no direct relation to superconductivity, being an energy scale of magnetic origin (the 'pseudogap' idea introduced by Friedel, as quoted in ref. 8). This would be in opposition to the claim by Miyakawa *et al.*¹⁵ that the pseudogap becomes, below T_c , the full BCS d -wave gap, with no other energy scale involved in the condensed state. It is remarkable, however, that both energy scales appear to join smoothly in the overdoped regime. One may also be concerned by the fact that Δ_p

Table 1 Doping dependence of the strong coupling ratio

Doping level	Arpes*	Tunnelling	Andreev reflection	$\lambda(T)$	Raman scattering
Optimally doped					
BSCCO	9 (ref. 3)	5 (refs 4, 15)		6 (ref. 20)	6 (ref. 22)
YBCO		8 (ref. 11)	5 (ref. 12)	5 (ref. 20)	
LSCO		7.5 (refs 17, 18)	5 (ref. 17)		5 (ref. 25)
HgBa ₂ CaCu ₂ O _{6+δ}					5 (ref. 26)
Overdoped					
BSCCO (60 K)		6–9 (refs 4, 15)			6 (ref. 22)
Underdoped					
BSCCO (80 K)	11–13 (ref. 27)	12 (refs 4, 15)	6.2 (ref. 16)		
BSCCO (70 K)		14 (ref. 15)			5 (ref. 22)
YBCO (60–65 K)		20 (ref. 3)	4.6 (ref. 14)	5 (ref. 21) 4 (ref. 20)	
LSCO (15 K)			4.5 (ref. 17)		

The values of the strong coupling ratio ($2\Delta/kT_c$) shown here were obtained using the indicated methods of gap measurement (see text for details). The strong coupling ratio ranges from 6 to 20 for Δ_p (ARPES, tunnelling), and from 4 to 6 for Δ_c (Andreev reflection, $\lambda(T)$, Raman scattering).

* Angle-resolved photoemission spectroscopy.

values have been obtained by surface-sensitive measurements, while Δ_c values have been obtained from both surface (Andreev) and bulk measurements ($\lambda(T)$ and Raman), and may therefore be more reliable. Yet the persistence of Δ_p above T_c in the underdoped regime is in good agreement with a decrease of the spin susceptibility below a temperature $T^* > T_c$, which is a bulk measurement¹.

A related question is that of the mechanism leading to pair formation. A non-mean-field behaviour may just indicate the presence of strong phase fluctuations, and not necessarily the presence of preformed pairs in the Mott sense²³. Both local pairs and the stripe model may explain the continuous increase of Δ_p in the underdoped regime. But if the existence of two energy scales was simply due to fluctuation effects, one would expect Δ_p to saturate in the underdoped regime²⁴. It is also possible that Δ_p might be unrelated to a pairing amplitude.

In any case, I consider that the existence of two energy scales is well established in the underdoped regime, with the scales converging in the overdoped regime. I have assigned the lower scale, obtained from Andreev reflection experiments, to coherence properties of the condensate; the higher scale reflects the properties of single-particle excitations. □

Note added in proof: I have learned of theoretical work on the superfluid density and the excitation gap in a BCS–Bose–Einstein crossover scenario²⁸.

Received 21 July; accepted 23 November 1998.

- Alloul, H., Ohno, T. & Mendels, P. Y. NMR evidence for a Fermi liquid behavior in YBa₂Cu₃O_{6+x}. *Phys. Rev. Lett.* **63**, 1700–1703 (1989).
- Ding, H. *et al.* Spectroscopic evidence for a pseudogap in the normal state of underdoped high- T_c superconductors. *Nature* **382**, 51–54 (1996).
- Racah, D. & Deutscher, G. Properties of normal metal/dielectric/high T_c junctions obtained by in-situ oxidation. *Physica C* **263**, 218–224 (1996).
- Renner, C. *et al.* Pseudogap precursor of the superconducting gap in under- and overdoped Bi₂Sr₂CaCu₂O_{8-δ}. *Phys. Rev. Lett.* **80**, 149–152 (1998).
- Griffin, A., Snook, D. W. & Stringari, S. (eds) *Bose-Einstein Condensation* (Cambridge Univ. Press, 1995).
- Emery, V. J., Kivelson, S. A. & Zachar, O. Spin-gap proximity effect mechanism of high temperature superconductivity. *Phys. Rev. B* **56**, 6120–6147 (1997).
- Deutscher, G. Superconducting glass and related properties. *Physica C* **153–155**, 15–20 (1988).
- Pines, D. The spin fluctuation model for high temperature superconductivity: progress and prospects. In *The Gap Symmetry and Fluctuations in High Temperature Superconductors* (eds Bok, J., Deutscher, G., Pavuna, D. & Wolf, S.) (NATO ASI ser., Plenum, 1998).
- Uemura, Y. J. Bose Einstein to BCS cross-over picture for high T_c cuprates. *Physica C* **282–287**, 194–197 (1997).
- Zhao, G.-m., Hunt, M. B., Keller, H. & Müller, K. A. Evidence for polaronic supercarriers in the copper oxide superconductors La_{2-x}Sr_xCuO₄. *Nature* **385**, 236–239 (1996).
- Maggio-Aprile, I. *et al.* Direct vortex imaging and tunneling spectroscopy of flux lines in YBa₂Cu₃O_{7-δ}. *Phys. Rev. Lett.* **75**, 2754–2757 (1997).
- Hass, N. *et al.* Sharp gap edge and determination of the Fermi velocity in YBa₂Cu₃O_{7-δ} by point contact spectroscopy. *J. Supercond.* **5**, 191–194 (1992).
- Tanaka, Y. in *Coherence in High Temperature Superconductors* (eds Deutscher, G. & Revcolveschi, A.) 393–413 (World Scientific, Singapore, 1996).

- Yagil, Y. *et al.* Experimental evidence for strong electron-phonon coupling to selected phonon modes in point contact spectroscopy. *Physica C* **250**, 59–66 (1995).
- Miyakawa, N. *et al.* Strong dependence of the superconducting gap on oxygen doping from tunneling measurements on Bi₂Sr₂CaCu₂O_{8-δ}. *Phys. Rev. Lett.* **80**, 157–160 (1998).
- Sinha, S. & Ng, K. W. Zero bias conductance peak enhancement in Bi₂Sr₂CaCu₂O₈/Pb tunneling junctions. *Phys. Rev. Lett.* **80**, 1296–1299 (1998).
- Achraf, N. *et al.* in *Coherence in High Temperature Superconductors* (eds Deutscher, G. & Revcolveschi, A.) 428–442 (World Scientific, Singapore, 1996).
- Ekino, T. *et al.* Superconducting energy gap of La_{1.85}Sr_{0.15}CuO₄ single crystals from breakjunction tunneling. *Physica C* **263**, 249–256 (1996).
- Kohen, A. *Transition from s-wave to d-wave Gap Symmetry in La_{2-x}Sr_xCuO₄*. Thesis, Tel Aviv Univ. (1997).
- Panapogoulos, C., Cooper, J. R. & Xiang, T. Systematic behavior of the in-plane penetration depth in d-wave superconductors. *Phys. Rev. B* **57**, 13422–13425 (1998).
- Bonn, D. A. *et al.* Comparison of the influence of Ni and Zn impurities on the electromagnetic properties of YBa₂Cu₃O_{6.95}. *Phys. Rev. B* **50**, 4051–4063 (1995).
- Nemetschek, R. *et al.* Pseudogap and superconducting gap in the electronic Raman spectra of underdoped cuprates. *Phys. Rev. Lett.* **78**, 4837–4840 (1997).
- Mott, N. V. High temperature superconductors above T_c . *Physica A* **168**, 221–228 (1990).
- Lee, P. A. & Wen, X. G. Unusual superconducting state of underdoped cuprates. *Phys. Rev. Lett.* **21**, 4111–4114 (1997).
- Chen, X. K. *et al.* Investigation of the superconducting gap in La_{2-x}Sr_xCuO₄ by Raman scattering. *Phys. Rev. Lett.* **73**, 3290–3293 (1994).
- Scuto, A. *et al.* Nodes of the superconducting gap probed by electronic Raman scattering in HgBa₂CaCu₂O_{6+δ}. *Europhys. Lett.* **39**, 207–210 (1997).
- Harris, J. M. *et al.* Anomalous superconducting gap size versus T_c behavior in underdoped Bi₂Sr₂Ca_{1-x}Dy_xCu₂O_{8+δ}. *Phys. Rev. B* **54**, 15665–15668 (1996).
- Chen, Q. *et al.* *Phys. Rev. Lett.* **81**, 4708 (1998).

Acknowledgements. I thank J. Friedel, P. Nozieres, R. Maynard and B. K. Chakraverty for discussions on the two-energy scale issue, and V. Emery for detailed comments on the manuscript. This work was supported in part by the Israel Science Foundation, by the Heinrich Hertz Minerva Center for Superconductivity, by the US Navy, and by the Oren Family Chair for Experimental Solid State Physics.

Correspondence and requests for materials should be addressed to the author.

High-temperature weak ferromagnetism in a low-density free-electron gas

D. P. Young*, D. Hall*, M. E. Torelli*, Z. Fisk*†, J. L. Sarrao†, J. D. Thompson†, H.-R. Ott‡, S. B. Oseroff§, R. G. Goodrich|| & R. Zysler¶

* NIMFL, Florida State University, Tallahassee, Florida 32306, USA

† Material Science and Technology Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87454, USA

‡ Laboratorium für Festkörperphysik, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

§ Department of Physics, San Diego State University, San Diego, California 92182, USA

|| Department of Physics, Louisiana State University, Baton Rouge, Louisiana 70803, USA

¶ Centro Atomico Bariloche, Bariloche RN 8400, Argentina

The magnetic properties of the ground state of a low-density free-electron gas in three dimensions have been the subject of theoretical speculation and controversy for seven decades¹. Not only is this a difficult theoretical problem to solve, it is also a problem which has not hitherto been directly addressed experimentally. Here we report measurements on electron-doped calcium hexaboride (CaB₆) which, we argue, show that—at a density of 7×10^{19} electrons cm⁻³—the ground state is ferromagnetically polarized with a saturation moment of 0.07 μ_B per electron. Surprisingly, the magnetic ordering temperature of this itinerant ferromagnet is 600 K, of the order of the Fermi temperature of the electron gas.

The cubic hexaborides of the rare-earth elements have long attracted interest because of their wide variety of physical properties in spite of their simple crystallographic structures. These physical properties include very low work functions leading to the use of LaB₆ as a thermionic emitter^{2,3}, dense Kondo behaviour and electric quadrupole ordering in CeB₆ (ref. 4), Kondo insulating properties in SmB₆ (ref. 5), and low-carrier-density ferromagnetism in the local moment system EuB₆ (refs 6, 7).

The host material for the present experiments is the divalent alkaline-earth hexaboride CaB₆. The crystal structure of this material

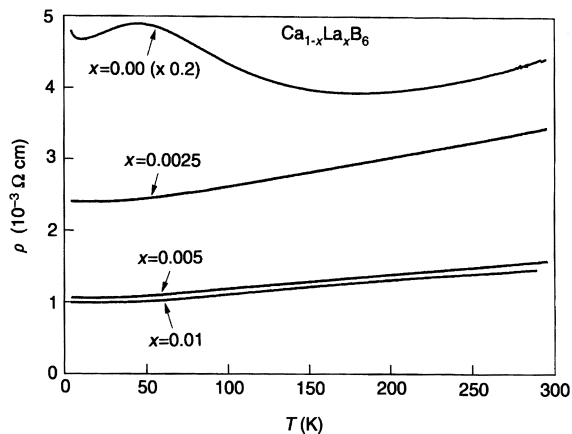


Figure 1 Temperature dependences of the electrical resistivities of La-doped CaB_6 . The resistivity for pure CaB_6 has been multiplied by 0.2.

can be thought of as a simple cubic CsCl-type arrangement of B_6 -octahedra and metal ions. The early electronic structure cluster calculations of Longuet-Higgins and Roberts⁸ found that the linked B_6 -network required 20 electrons for a 'closed shell' electronic configuration, indicating that the alkaline-earth hexaborides would be semiconductors. A study of the low-temperature properties of single crystals of SrB_6 , however, found a non-zero conductivity due to approximately 0.001 electrons per SrB_6 , and indications of the importance of electron-hole Coulomb effects⁹. More recent band structure calculations¹⁰ show that hexaborides with divalent cations should be semimetals, with a small direct overlap of a primarily boron-derived valence band with a primarily alkaline-earth-derived conduction band at the X-point in the Brillouin zone.

Our experiments, including measurements of the electrical resistivity, the magnetic susceptibility and the magnetization, were performed on single crystals of CaB_6 doped with trivalent La; these crystals were grown from stoichiometric amounts of the hexaboride components in molten Al. On similarly prepared single crystals of $\text{Sr}_{1-x}\text{Ce}_x\text{B}_6$, we have verified that Ce is incorporated stoichiometrically into the crystals by measuring the magnetic susceptibility and fitting the data to a Curie-Weiss law with an effective moment $\mu_{\text{eff}} = 2.54 \mu_B$ per trivalent Ce ion. It is reasonable to assume that the neighbour of Ce in the rare-earth sequence, La, is also incorporated stoichiometrically in SrB_6 and CaB_6 . X-ray diffraction has confirmed the cubic hexaboride structure in all cases.

Figure 1 shows the electrical resistivity data obtained between 5 K and 300 K from single crystals of $\text{Ca}_{1-x}\text{La}_x\text{B}_6$. Already for $x = 0.005$, we find a change to a metallic-like temperature dependence and a drop in resistivity by a factor of ~ 50 at low temperature, relative to the pure host ($x = 0$). A single band interpretation of the Hall constant obtained using the van der Pauw geometry for the $x = 0.005$ sample indicates electron-like carriers, with a density of 0.005 electrons per formula unit, as one naively expects for the trivalent La substitution. Our preliminary de Haas-van Alphen experiment on $x = 0.005$ La-doped CaB_6 along (100) reveals two orbits. Interpreting these as orthogonal extremal orbits of an ellipsoid gives an occupied volume (taking account of the ellipsoids at the equivalent X-points in the Brillouin zone) enclosing ~ 0.01 electrons for unpolarized electrons. This value is within experimental error of 0.005 electrons, given the small number of oscillations observed; the observation of two rather than four extremal orbits is consistent with the small hole pocket being filled, leaving only a small filling of the conduction band.

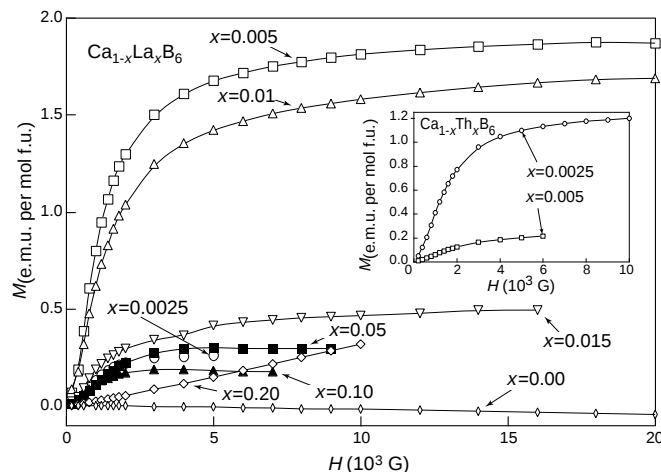


Figure 2 Magnetization per mol of compound versus applied field for La-doped CaB_6 , measured at 5 K (f.u., formula unit). The data for $x = 0$ were taken at 10 K. Inset, magnetization versus applied field at 5 K for $\text{Ca}_{0.9975}\text{Th}_{0.0025}\text{B}_6$ and $\text{Ca}_{0.995}\text{Th}_{0.005}\text{B}_6$. Magnetizations have been measured after cooling in zero field, using a Quantum Design SQUID magnetometer.

The unusual aspect of these La-doped borides is seen in Fig. 2, where we plot magnetization versus field at 5 K for a range of values of x , measured on single crystals of $\text{Ca}_{1-x}\text{La}_x\text{B}_6$. A weak ferromagnetic moment is evident, its magnitude peaking near $x = 0.005$ at $0.07 \mu_B$ per La, with no moment found in crystals with composition $\text{Ca}_{0.80}\text{La}_{0.20}\text{B}_6$ (Fig. 3); hysteresis loops for the $x = 0.005$ material are shown in Fig. 3 inset. We find nearly identical magnetic effects with La-doping of CaB_6 , SrB_6 and BaB_6 . In all cases, the maximum moment is found at $x = 0.005$. That this moment is a function of the carrier concentration is supported by data for $\text{Ca}_{0.9975}\text{Th}_{0.0025}\text{B}_6$ and $\text{Ca}_{0.995}\text{Th}_{0.005}\text{B}_6$ (Fig. 2 inset): here the moment is maximum at $x = 0.0025$. Because Th is incorporated in a tetravalent configuration, it will contribute one more electron than trivalent La, and hence if the moment is a function of carrier count, compounds with $\text{Th}_{0.0025}$ and $\text{La}_{0.005}$ should have the same moment. In Fig. 4, we show the temperature dependence of the magnetic moment of $\text{Ca}_{0.995}\text{La}_{0.005}\text{B}_6$ in a fixed field of 0.1 T as a function of temperature: the data show loss of magnetization near the Curie temperature, T_C , of approximately 600 K.

The essential point is to determine that the observed weak ferromagnetism is an intrinsic effect and not of some extrinsic origin. The magnitude of the ordered moment is ~ 1 e.m.u. per mol hexaboride for $x = 0.005$ in CaB_6 , and is ~ 2 e.m.u. per mol for $\text{Sr}_{0.995}\text{La}_{0.005}\text{B}_6$. We find consistently that the moment per mole is at a maximum near $x = 0.005$ in both SrB_6 and CaB_6 . The same systematic study has not yet been made for BaB_6 . We find similarly that Ce- and Sm-doped SrB_6 samples have an ordered moment which peaks at $x = 0.005$. In addition, however, we recognize the expected paramagnetic background due to the local magnetic moment due to the f-electrons of these atoms, which is absent for La-doped material. These data indicate that the doping does not produce the moment through trace rare-earth impurities carried by the high-purity La used for the doping. For the pure alkaline-earth hexaborides, we find that a moment is present sometimes, and that this varies from crystal to crystal. This moment is generally distinctly smaller than that found at $x = 0.005$, usually by at least an order of magnitude. We also find a variation in the temperature dependence of the electrical resistivity of pure SrB_6 and CaB_6 from crystal to crystal; crystals showing weak ferromagnetism have a more metallic temperature dependence of the electrical resistivity. Doping divalent hexaborides with other alkaline earths also produces weak ferromagnetism, for instance in $\text{Ca}_{0.995}\text{Ba}_{0.005}\text{B}_6$. We can understand this variable behaviour of the crystals with no

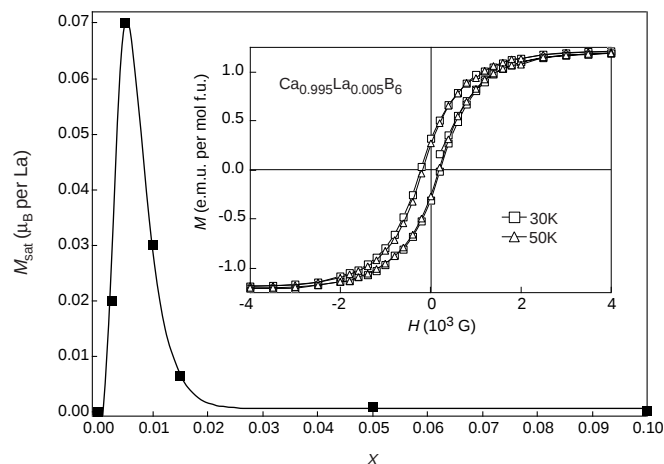


Figure 3 Saturation moment at $T = 5$ K per mol La as a function of La concentration in CaB_6 . The line is a guide to the eye. Inset, hysteresis loops at 30 K and 50 K for $\text{Ca}_{0.995}\text{La}_{0.005}\text{B}_6$.

carrier doping on the basis of the band structure of the alkaline-earth hexaborides. Calculations¹⁰ show that the details of the overlap of valence and conduction bands at the X-point of the Brillouin zone depend sensitively on the crystallographic parameter fixing the location of the borons in the unit cell; this parameter determines the relative length of the inter- and intra-octahedral boron–boron bonds. Even small changes in this parameter in the calculations alter the behaviour from insulating to metallic. So we might expect vacancies and foreign-atom additions to alter significantly the properties of the divalent hexaborides for very small dopings, such as seen in materials with similar band structure features (for example, grey tin and bismuth).

Two suggestions for the origin of the weak ferromagnetism which we have observed are (1) ordered defect moments and (2) ferromagnetic polarization of the low-density electron gas. Because no obvious source for a strong coupling giving rise to a Curie temperature as high as $T_c = 600$ K presents itself, the coupling of magnetic moments localized on the La or other unspecified impurities on this scale seems rather implausible. A more likely candidate for the origin of the magnetic polarization emerges from studies of electronic correlations in the low-density electron gas, such as those of Ceperley and Alder¹¹. This is a topic of theoretical speculation with a long history, going back to Bloch and Wigner¹. The study of Ceperley and Alder is a $T = 0$ K computation, comparing unpolarized and completely polarized states of the electron gas, with ferromagnetism showing up for values of r_s of the order $80 a_B$. (Here r_s is the radius of the sphere containing one conduction electron; a_B the Bohr radius.) Later calculations have lowered this value to $\sim 20 a_B$ (ref. 12). For $x = 0.005$, we compute $r_s = 15.0 \text{ \AA} = 28.4 a_B$, using the Bohr radius for the free electron. A recent calculation by Ortiz, Harris and Ballone¹³ treating partially spin-polarized states of the low-density electron gas has, in fact, found evidence that near $r_s = 30 a_B$ the stable state is one with ferromagnetic polarization of the order of 10%; this is essentially our experimental finding of an ordered moment of $\sim 0.07 \mu_B$ per carrier at $x = 0.005$. The natural energy scale here is the Fermi energy, E_F ; for free electrons we have $E_F = 0.062 \text{ eV} = 720 \text{ K}$ for $x = 0.005$ in CaB_6 , of the order of the observed Curie temperature.

The ferromagnetic ground state of a dilute, three-dimensional electron gas has not previously been reported experimentally. But such a ground state seems to provide a possible description of the weak ferromagnetism reported here, although the temperature scale of the phenomenon is unexpected. Detailed calculations appropriate to the lattice case at finite temperature are clearly needed, as

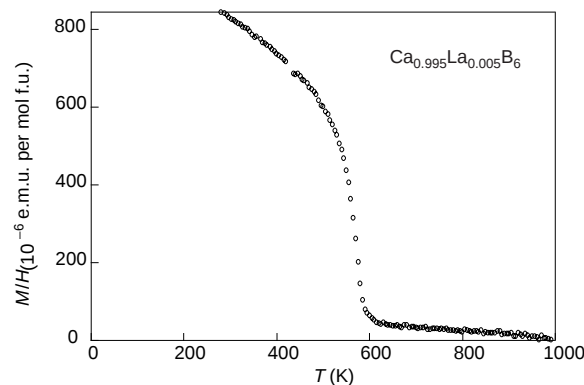


Figure 4 Magnetization of $\text{Ca}_{0.995}\text{La}_{0.005}\text{B}_6$ in a fixed applied field of 0.1 T as a function of temperature. These data were measured using a Faraday balance magnetometer.

is much further experimental elaboration of the details of this ferromagnetism. □

Received 14 July; accepted 4 November 1998.

- Herring, C. *Magnetism* Vol. IV (eds Rado, G. T. & Suhl, H.) Ch. 2 (Academic, New York, 1966).
- Schmidt, P. H. *et al.* Anisotropy of thermionic electron emission values for LaB_6 single-crystal emitter cathodes. *Appl. Phys. Lett.* **29**, 400–410 (1976).
- Lafferty, J. M. *J. Appl. Phys.* **22**, 299–301 (1951).
- Erkelens, W. A. C. *et al.* Neutron scattering study of the antiferroquadrupolar ordering in CeB_6 and $\text{Ce}_{0.75}\text{La}_{0.25}\text{B}_6$. *J. Magn. Mater.* **63/64**, 61–63 (1987).
- Nyhus, P., Cooper, S. L., Fisk, Z. & Sarrao, J. L. Low-energy excitations of the correlation-gap insulator SmB_6 . *Phys. Rev. B* **55**, 12488–12496 (1997).
- Süllow, S. *et al.* Structure and magnetic order of EuB_6 . *Phys. Rev. B* **57**, 5860–5869 (1998).
- Degiori, L., Felder, E., Ott, H. R., Sarrao, J. L. & Fisk, Z. Low-temperature anomalies and ferromagnetism in EuB_6 . *Phys. Rev. Lett.* **79**, 5134–5137 (1997).
- Longuet-Higgins, H. C. & Roberts, M. de V. The electronic structure of the borides MB_6 . *Proc. R. Soc. Lond. A* **224**, 336–347 (1954).
- Ott, H. R. *et al.* Structure and electronic properties of SrB_6 . *Z. Phys. B* **102**, 337–345 (1997).
- Massida, S., Continenza, A., de Pascale, T. M. & Monnier, R. Z. Electronic structure of divalent hexaborides. *Z. Phys. B* **102**, 83–89 (1997).
- Ceperley, D. M. & Alder, B. J. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.* **45**, 566–569 (1980).
- Alder, B. J., Ceperley, D. M. & Pollock, E. L. *Int. J. Quantum Chem. Symp.* **16**, 49–61 (1982).
- Ortiz, G., Harris, M. & Ballone, P. cond-mat/9810126.

Acknowledgements. Z.F. thanks E. Abrahams, G. Aeppli, L. Gor'kov, P. Kumar, E. Manousakis, G. Ortiz and D. Pines for discussions. Work at NHMFL was supported by the NSF; work at Los Alamos was done under the auspices of the US Department of Energy; and the work at ETH Zürich was supported by the Schweizerische Nationalfonds zur Förderung der wissenschaftlichen Forschung.

Correspondence and requests for materials should be addressed to Z.F. (e-mail: fisk@magnet.fsu.edu).

Improved quantum efficiency for electroluminescence in semiconducting polymers

Yong Cao, Ian D. Parker, Gang Yu, Chi Zhang & Alan J. Heeger

UNIAx Corporation, 6780 Cortona Drive, Santa Barbara, California 93117-3022, USA

Some conjugated polymers have luminescence properties that are potentially useful for applications such as light-emitting diodes, whose performance is ultimately limited by the maximum quantum efficiency theoretically attainable for electroluminescence^{1,2}. If the lowest-energy excited states are strongly bound excitons (electron–hole pairs in singlet or triplet spin states), this theoretical upper limit is only 25% of the corresponding quantum

efficiency for photoluminescence: an electron in the π^* -band and a hole (or missing electron) in the π -band can form a triplet with spin multiplicity of three, or a singlet with spin multiplicity of one, but only the singlet will decay radiatively³. But if the electron-hole binding energy is sufficiently weak, the ratio of the maximum quantum efficiencies for electroluminescence and photoluminescence can theoretically approach unity. Here we report a value of $\sim 50\%$ for the ratio of these efficiencies (electroluminescence:photoluminescence) in polymer light-emitting diodes, attained by blending electron transport materials with the conjugated polymer to improve the injection of electrons. This value significantly exceeds the theoretical limit for strongly bound singlet and triplet excitons, assuming they comprise the lowest-energy excited states. Our results imply that the exciton binding energy is weak, or that singlet bound states are formed with higher probability than triplets.

The devices for the electroluminescence (EL) and photoluminescence (PL) measurements were fabricated in the thin sandwich configuration: anode/polymer/cathode. The electronic structure of alkoxy derivatives of poly(*p*-phenylene vinylene), PPV, has been studied via internal field emission⁴, internal photoemission⁵, electrochemical doping (cyclic voltammetric spectroscopy)⁶ and photoelectron spectroscopy⁷. The data indicate that the π^* -band and the top of the π -band are at ~ 3 eV and ~ 5 eV respectively, with respect to the vacuum. Thus, relatively good hole and electron injection can be achieved by using transparent indium-tin oxide (ITO) as the anode and calcium (Ca) or barium (Ba) as the cathode^{2,4}. However, hole injection is very sensitive to the quality of the ITO. We have found that a layer (~ 30 nm) of conducting polyaniline⁸, polypyrrole⁹, or poly(ethylene-dioxythiophene)¹⁰ provides excellent, reproducible hole-injecting contacts⁸⁻¹⁰.

Figure 1 shows the external quantum efficiency for electroluminescence, $QE_{\text{ext}}(\text{EL})$, as a function of film thickness for two closely related PPV derivatives; OC1C10-PPV and MEH-PPV. The oscillatory dependence on the thickness (d) of the polymer layer arises from the proximity of the metallic mirror electrode; the emitting oscillator interacts with a virtual image oscillator "behind the mirror"^{11,12}. Because the radiation from the emitter and the retarded radiation from the image oscillator interfere, the PL decay rate is an oscillatory function of the distance from the mirror. Consequently, for a thin film, the quantum yields for PL and EL are oscillatory functions of d . When the average distance from the mirror to the oscillator is too small ($d \ll 100$ nm), the losses from the metallic electrode quench the luminescence.

Blom *et al.* suggested that a high density of electron traps inhibits electron transport and prevents the achievement of balanced injection¹³⁻¹⁵. In an attempt to address this problem, we have studied the effect of blending electron transport materials into the conjugated polymer. The best results were obtained with (2-(4-biphenyl)-5-(4-tert-butylphenyl)1,3,4-oxadiazole, Bu-PBD, which was blended into the EL polymer at various concentrations. The butylphenyl substitution in Bu-PBD leads to good compatibility with the alkoxy-PPVs.

The light-emitting diodes (LEDs) were sealed with a glass cover that enables extended lifetimes at elevated temperatures. The external quantum efficiencies for PL and EL were directly compared for the OC1C10-PPV plus Bu-PBD blends in the LED configuration. For measurements of the EL efficiency, the external emission, integrated over all directions, was collected with a calibrated integrating sphere (at room temperature). PL efficiency measurements on the identical devices used the same integrating sphere¹⁶. For PL measurements, the excitation source was a monochromatic

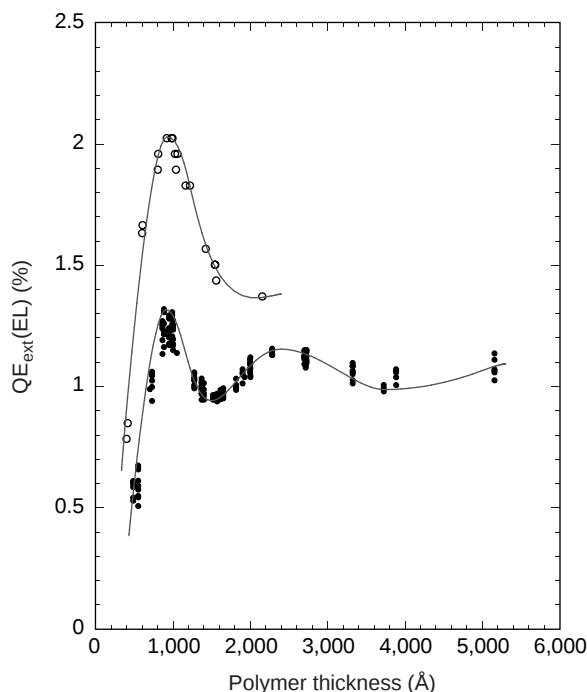


Figure 1 External electroluminescence quantum efficiency, $QE_{\text{ext}}(\text{EL})$, as a function of film thickness. Data are shown for two closely related alkoxy derivatives of poly(*p*-phenylene vinylene): OC1C10-PPV and MEH-PPV. The data from OC1C10-PPV [poly(2-(3,7-dimethyloctyloxy)-5-(2'-methoxy-1,4-phenylene vinylene))] form the upper curve (open circles); the data from MEH-PPV [poly(2-methoxy-5-(2'-ethylhexyloxy)-1,4-phenylene vinylene)] form the lower curve (filled circles). $QE_{\text{ext}}(\text{EL})$ is given in per cent; $1\% = 10^{-2}$ photons per carrier. The solid curves are guides to the eye.

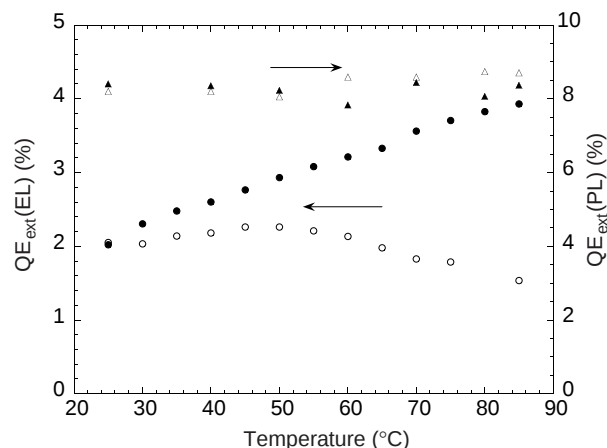


Figure 2 Temperature dependence of $QE_{\text{ext}}(\text{EL})$ and $QE_{\text{ext}}(\text{PL})$. Filled circles, $QE_{\text{ext}}(\text{EL})$ of a device fabricated from OC1C10-PPV plus 20% Bu-PBD. Data from an identical device fabricated with pure OC1C10-PPV (without Bu-PBD) are shown for comparison (open circles). The EL data were obtained at 6.7 mA cm^{-2} . Temperature dependences of the PL are also shown: pure OC1C10-PPV, open triangles; OC1C10-PPV plus 20% Bu-PBD, filled triangles.

laser beam (wavelength 488 nm) incident perpendicular to the plane of the device from the glass/ITO side. Extension of the external quantum efficiency measurements, PL and EL, to elevated temperatures used a photodiode that was carefully calibrated by the integrating sphere. By making the PL and EL measurements with the same device, the PL and EL data are compared in the same film and with the same electrodes. Using the device configuration simplifies the determination of the ratio of quantum efficiencies. Initially ignoring possible differences in recombination zones for PL and EL, the loss factors which contribute to both PL and EL divide out; both are subjected to quenching by proximity to the electrodes and to losses from internal reflection (and waveguiding), scattering and absorption. Thus, identical ratios are expected for the internal and external $QE(EL)/QE(PL)$.

Figure 2 shows the temperature dependence of the external quantum efficiency for electroluminescence, $QE_{ext}(EL)$, of a device fabricated from a blend of OC1C10-PPV containing 20% (by weight) Bu-PBD (film thickness 100 nm). The data from an identical device fabricated with pure OC1C10-PPV (without Bu-PBD) are shown for comparison. For devices with Bu-PBD, QE_{ext} increases with temperature; at 85 °C, $QE_{ext}(EL)$ is twice that obtained at room temperature (for devices made without Bu-PBD, $QE_{ext}(EL)$ decreases slightly with temperature). The increase in QE_{ext} is unambiguous; measurements were taken on more than 10 devices for each Bu-PBD concentration with excellent reproducibility. The increase with temperature is reversible. The increase is a function of the Bu-PBD content; the optimum is at ~20% Bu-PBD. At higher concentrations, black spots are observed, indicative of phase separation.

The increase in $QE_{ext}(EL)$ does not result from an increase in the external quantum efficiency for photoluminescence, $QE_{ext}(PL)$; as shown in Fig. 2, the latter is temperature independent between

room temperature and 85 °C in devices fabricated with and without the Bu-PBD. The temperature dependence of the operating voltage (for example, the voltage at 6.67 mA cm^{-2}) is the same for devices fabricated with and without Bu-PBD. Thus, the increase in QE_{ext} is not related to a change in mobility with temperature. This conclusion was confirmed by the observation that a bilayer device (an 8-nm layer of Bu-PBD deposited on top of OC1C10-PPV) showed a similar (though somewhat smaller) increase in QE_{ext} with temperature. Thus, the effect of the Bu-PBD is evidently to fine-tune the balance of the electron and hole injection. In this context, we note that the acceptor level in Bu-PBD is close in energy to the bottom of the π^* -band of the luminescence polymer¹⁷.

The PL and EL efficiencies were carefully checked on each of a series of devices fabricated with OC1C10-PPV films with different thickness (Fig. 3). The EL data are consistent with those in Fig. 1. The PL and EL efficiencies track one another as d is varied. We conclude from our results (Figs 2 and 3) that the external EL quantum efficiency for LEDs fabricated from OC1C10-PPV plus Bu-PBD (100-nm-thick) reaches $(4.0 \pm 0.2) \times 10^{-2}$ photons per electron (that is, $4.0(\pm 0.2)\%$) at 85 °C. The external PL quantum efficiency as obtained from OC1C10-PPV films, with and without the Bu-PBD, in identical devices is $8 \pm (0.8)\%$ at zero field. By using EL and PL data from the same devices (made with OC1C10-PPV plus 20% Bu-PBD) to minimize the error, we conclude that $QE_{ext}(EL)/QE_{ext}(PL) \approx 0.5$.

The data in Fig. 2 were taken at constant current density (6.67 mA cm^{-2}) with operating voltages of 3–4 V; that is, at electric fields of $(3\text{--}4) \times 10^5 \text{ V cm}^{-1}$. Previous measurements have demonstrated field-induced quenching of the PL; at 4 V, the PL quantum efficiency is reduced by ~10% (refs 18, 19). Thus, for the OC1C10-PPV plus Bu-PBD devices at 85 °C, $QE_{ext}(EL)/QE_{ext}(PL) = 0.5$ is a lower limit.

The quantum efficiencies for PL and EL are reduced through quenching by proximity to the metal and ITO electrodes. Quenching of the PL was directly measured by comparing PL efficiencies of a polymer film with $d = 100 \text{ nm}$ on glass (17.4%), on ITO/glass (12.3%) and on ITO/glass with a metal cathode (8.3%). In each case, the measurements were made in an integrating sphere with the polymer excited through the glass substrate. The data indicate approximately equal reductions in PL efficiency from the two electrodes. Because of the high density of electron traps, the recombination zone for EL is expected to be closer to the cathode, consistent with images of LEDs in the surface cell configuration²⁰. Greenham *et al.*^{12,16} showed that the luminescence is quenched within ~30–40 nm of the cathode. Quenching in the 30-nm-layer adjacent to the cathode was confirmed by analysis of the device characteristics^{13–15}. Because of the spatially shifted recombination zones, the EL efficiency falls more rapidly than the PL efficiency as the thickness is reduced below 100 nm (Fig. 3). Consequently, the conclusion that $QE_{ext}(EL)/QE_{ext}(PL) \approx 0.5$ at 85 °C is conservative; the EL efficiency suffers a greater reduction from electrode quenching than the PL efficiency.

Because of the total internal reflection only a fraction $(1/2n^2)$, where $n \approx 1.7$ is the index of refraction of the polymer) is emitted through the substrate²¹. That fraction of the emitted radiation which is waveguided is partially lost as a result of self-absorption in the polymer and, more importantly, as a result of losses in the metal cathode and the semitransparent anode. Consequently, the external quantum efficiencies are smaller than the internal quantum efficiencies. We find $QE_{ext}(PL) \approx 8\%$ in the device configuration in comparison with $QE_{int}(PL) \approx 15\text{--}20\%$, as obtained from direct measurements of the PL efficiency of OC1C10-PPV using the method of Greenham *et al.*¹⁶ Thus, the measured reduction factor is approximately 2–2.5, less than the predicted value ($2n^2 \approx 6$). We conclude that some of the light which is waveguided along the film/substrate emerges into the integrating sphere through a combination of surface scattering and emission from the edges.

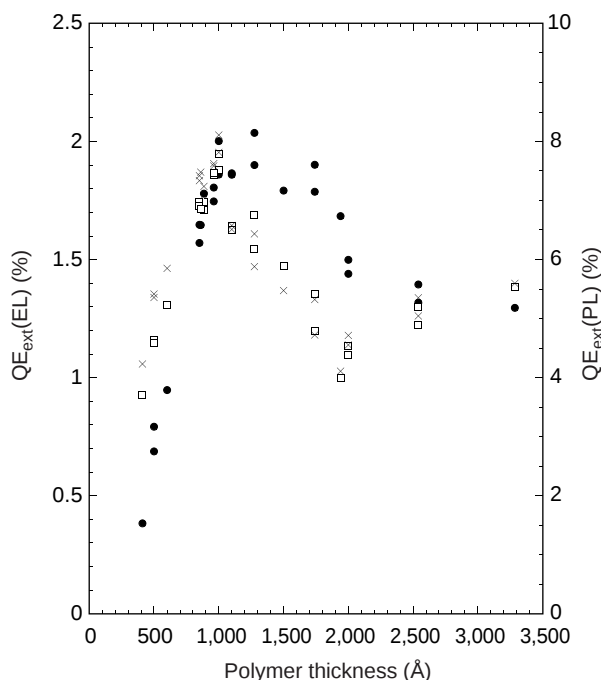


Figure 3 Comparison of external $QE_{ext}(EL)$, filled circles, with $QE_{ext}(PL)$, for devices with different luminescent polymer thickness. Open squares, PL efficiency as obtained when the incident light intensity was corrected by directly measuring the sum of unabsorbed and reflected light for each individual device outside the integrating sphere. Crosses, PL efficiency as obtained when the incident light intensity was corrected by measuring transmittance and reflectance of EL polymer films of different thickness on the glass substrate, following the procedure of Greenham *et al.*¹⁶

The external PL and EL quantum efficiencies of OC1C10-PPV blended with Bu-PBD have been quantitatively compared. The ratio of the EL and PL external efficiencies, $QE_{\text{ext}}(\text{EL})/QE_{\text{ext}}(\text{PL}) \approx 0.5$, is well beyond the theoretical limit for strongly bound singlet and triplet excitons. A summary and discussion of the magnitude of the exciton binding energy in semiconducting polymers has recently been published; values range from less than 0.1 eV to as high as 1 eV (ref. 22). Efficient triplet-to-singlet conversion (for example, by triplet-triplet annihilation) before non-radiative recombination (with subnanosecond decay time) is unlikely in hydrocarbon polymers where the spin-orbit coupling is weak. Thus, the high value for $QE(\text{EL})/QE(\text{PL})$ indicates either that the exciton binding energy is small or that the cross-section for an electron-hole pair to form a singlet bound state is significantly higher than the cross-section to form a triplet. Evidence of photogenerated triplet excitations has been reported (ref. 23 and references therein). However, the relevant timescale was in the millisecond regime or longer; hence, the observed triplets might be defect-stabilized.

Our results indicate that $QE(\text{EL})/QE(\text{PL})$ can be large in polymer LEDs. The goal is to achieve balanced injection with high electron and hole currents in materials with large $QE(\text{EL})$. As high PL efficiencies (>60%) have been demonstrated²⁴, the achievement of efficient EL emission from polymer LEDs should be possible. □

Received 15 April; accepted 12 November 1998.

- Burroughes, J. H. *et al.* Light-emitting diodes based on conjugated polymers. *Nature* **347**, 539–541 (1990).
- Braun, D. & Heeger, A. J. Visible light emission from semiconducting polymer diodes. *Appl. Phys. Lett.* **58**, 1982–1984 (1991).
- Bradley, D. D. C. Conjugated polymer electroluminescence. *Synth. Met.* **54**, 401–415 (1993).
- Parker, I. D. Carrier tunneling and device characteristics in polymer light-emitting diodes. *J. Appl. Phys.* **75**, 1656–1666 (1994).
- Campbell, I. H., Hagler, T. W., Smith, D. L. & Ferraris, J. P. Direct measurement of conjugated polymer electronic excitation energies using metal/polymer/metal structures. *Phys. Rev. Lett.* **76**, 1900–1903 (1996).
- Richter, M. M., Fan, F., Klavetter, F., Heeger, A. J. & Bard, A. Electrochemistry and electrogenerated chemiluminescence of films of the conjugated polymer 4-methoxy-(2-ethylhexyloxy)-2,5-polyphenylenevinylene. *Chem. Phys. Lett.* **226**, 115–120 (1994).
- Brom, P., Birgersson, J., Johansson, N., Logdland, M. & Salaneck, W. R. Calcium electrodes in polymer LEDs. *Synth. Met.* **74**, 179–181 (1995).
- Yang, Y., Westerweel, E., Zhang, C., Smith, P. & Heeger, A. J. Enhanced performance of polymer light-emitting diodes using high-surface area polyaniline network electrodes. *J. Appl. Phys.* **77**, 694–698 (1995).
- Gao, J., Heeger, A. J., Lee, J. Y. & Kim, C. Y. Soluble polypyrrole as the transparent anode in polymer light-emitting diodes. *Synth. Met.* **82**, 221–223 (1996).
- Cao, Y., Yu, G., Zhang, C., Menon, R. & Heeger, A. J. Polymer light-emitting diodes with polyethylene dioxathiophene-polystyrene sulfonate as the transparent anode. *Synth. Met.* **87**, 171–174 (1997).
- Kuhn, H. Classical aspects of energy transfer in molecular systems. *J. Chem. Phys.* **53**, 101–108 (1970).
- Greenham, N. C. Understanding the limits to performance in electronic devices based on PPV. *Bull. Am. Phys. Soc.* **43** (1), 14 (1998).
- Blom, P. W. M., de Jong, M. J. M. & Vleggaar, J. J. M. Electron and hole transport in poly(p-phenylene vinylene). *Appl. Phys. Lett.* **68**, 3308–3310 (1996).
- Blom, P. W. M., de Jong, M. J. M. & Munster, M. G. Electric-field and temperature dependence of the hole mobility in poly(p-phenylene vinylene). *Phys. Rev. B* **55**, 656–659 (1997).
- Blom, P. W. M., de Jong, M. J. M. & Breedijk, S. Temperature dependence electron-hole recombination in polymer light-emitting diodes. *Appl. Phys. Lett.* **71**, 930–932 (1997).
- Greenham, N. C. *et al.* Measurement of absolute photoluminescence quantum efficiencies in conjugated polymer. *Chem. Phys. Lett.* **241**, 89–96 (1995).
- Schmidt, A., Anderson, M. L. & Armstrong, N. R. Electronic states of vapor deposited electron and hole transport agents and luminescent materials for light-emitting diodes. *J. Appl. Phys.* **78**, 5619–5625 (1995).
- Keresting, R. *et al.* Ultrafast field-induced dissociation of excitons in conjugated polymers. *Phys. Rev. Lett.* **73**, 1440–1443 (1994).
- Moses, D. *et al.* Mechanism of carrier generation in poly(phenylene vinylene): Transient photoconductivity and photoluminescence at high fields. *Phys. Rev. B* **54**, 4748–4754 (1996).
- Greenham, N. C., Friend, R. H. & Bradley, D. D. C. Angular dependence of the emission from a conjugated polymer light-emitting diode: implications for efficiency calculations. *Adv. Mater.* **6**, 491–494 (1994).
- Lemmer, U. *et al.* Electroluminescence from poly(phenylene vinylene) in a planar metal-polymer-metal structure. *Appl. Phys. Lett.* **68**, 3007–3009 (1996).
- Sariciftci, N. S. (ed.) *Primary Photoexcitations in Conjugated Polymers: Molecular Excitons vs Semiconductor Band Model* (World Scientific, Singapore, 1997).
- Smilowitz, L. & Heeger, A. J. Photoinduced absorption from triplet excitations in poly(2-methoxy-5-(2'-ethylhexyloxy)-p-phenylene vinylene) oriented by gel processing in polyethylene. *Synth. Met.* **48**, 193–196 (1992).
- Hide, F. *et al.* Semiconducting polymers: a new class of solid state laser materials. *Science* **273**, 1833–1836 (1996).

Acknowledgements. Research at UNIAx was partially supported by the Office of Naval Research.

Correspondence and requests for materials should be addressed to A.J.H.

Hybrid hydrogels assembled from synthetic polymers and coiled-coil protein domains

Chun Wang*, Russell J. Stewart* & Jindřich Kopeček*†

Departments of * Bioengineering and † Pharmaceuticals and Pharmaceutical Chemistry, University of Utah, Salt Lake City, Utah 84112, USA

Stimuli-sensitive polymer hydrogels, which swell or shrink in response to changes in the environmental conditions, have been extensively investigated and used as 'smart' biomaterials and drug-delivery systems^{1,2}. Most of these responsive hydrogels are prepared from a limited number of synthetic polymers and their derivatives, such as copolymers of (meth)acrylic acid, acrylamide and *N*-isopropyl acrylamide^{3–12}. Water-soluble synthetic polymers have also been crosslinked with molecules of biological origin, such as oligopeptides¹³ and oligodeoxyribonucleotides¹⁴, or with intact native proteins¹⁵. Very often there are several factors influencing the relationship between structure and properties in these systems, making it difficult to engineer hydrogels with specified responses to particular stimuli. Here we report a hybrid hydrogel system assembled from water-soluble synthetic polymers and a well-defined protein-folding motif, the coiled coil. These hydrogels undergo temperature-induced collapse owing to the cooperative conformational transition of the coiled-coil protein domain. This system shows that well-characterized water-soluble synthetic polymers can be combined with well-defined folding motifs of proteins in hydrogels with engineered volume-change properties^{16,17}.

The coiled coil¹⁸, a left-handed superhelix of two or more right-handed α -helices, has been identified in proteins ranging from

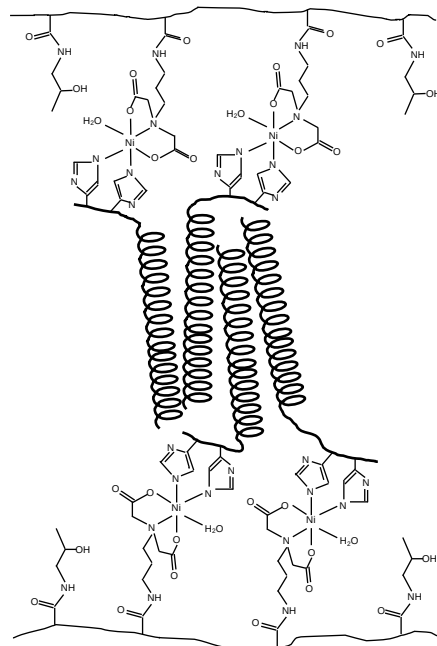


Figure 1 Structural representation of the hybrid hydrogel primary chains and the attachment of His-tagged coiled-coil proteins. Poly(HPMA-co-DAMA) is shown here as the primary chains. The pendant iminodiacetate groups form complexes with transition-metal ions, such as Ni^{2+} , to which the terminal histidine residues of the coiled coils are attached. A tetrameric coiled coil (not drawn to scale), consisting of two parallel dimers associating in an antiparallel fashion, is shown here as an example of many of the possible conformations.

muscle proteins to DNA transcription factors. It has the characteristic amino-acid heptad repeating units designated as 'a' to 'g' with hydrophobic residues at positions 'a' and 'd', and polar residues at other positions. Many native and *de novo* designed coiled coils can undergo conformational transition in response to temperature, pH, ionic strength, solvent and so forth. Dramatic changes in coiled-coil stability, conformation, and responsiveness towards the external environment can be caused by minor alteration of the primary structure; such alterations can occur through natural evolution¹⁹, experimental mutagenesis²⁰ or *de novo* design^{21,22}. We hypothesized that by using coiled coils with engineered sequences as crosslinks for synthetic polymer chains, we would be able to prepare hybrid hydrogels (Fig. 1) with predetermined stimuli sensitivity reflecting the structure and properties of coiled coils.

For the primary chains of our hybrid hydrogels (Fig. 1), we prepared a linear hydrophilic copolymer of *N*-(2-hydroxypropyl)-methacrylamide (HPMA)¹³, and a metal-chelating monomer *N*-(*N'*,*N'*-dicarboxymethylaminopropyl)methacrylamide (DAMA), by radical copolymerization. A metal complex is formed (Fig. 1)

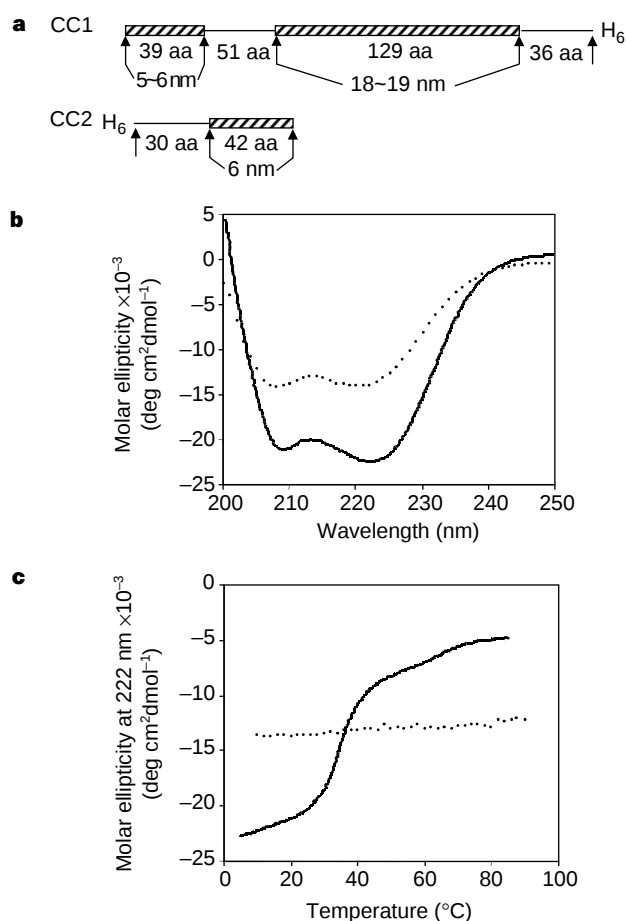


Figure 2 The coiled coils and characterization by circular dichroism. **a**, The block diagrams of the two proteins (CC1 and CC2). The amino-acid (aa) sequence of the EK42 domain is (VSSLESK)₆. The region separating the EK42 domain and the His tag is GMASMTGGQQMGRDLYDDDDKDP. The numbers of amino-acid residues in the coiled-coil and non-coiled-coil domains are given. The length of the coiled-coil domains was estimated according to ref. 28. **b**, CD spectra of the two proteins (solid line, CC1; dotted line, CC2) in PBS at 25 °C. The ratios of molar ellipticity minima at 222 and 208 nm were close to 1.0, typically found for α -helical coiled coils²⁷. **c**, Thermal unfolding of both proteins (solid line: CC1; dotted line: CC2) in PBS as monitored by the change in molar ellipticity at 222 nm wavelength.

by the pendant metal-chelating ligand—iminodiacetate (IDA)-Ni²⁺ and the terminal histidine residues (His tag) of the coiled coils. Similar schemes have been used for purifying recombinant proteins²³ and immobilizing them on interfaces^{24,25}. Here it serves as a simple, convenient way of anchoring genetically engineered His-tagged coiled coils to the hydrogel primary chains.

Two His-tagged coiled coils were used to assemble hybrid hydrogels. The first, designated CC1, was a natural protein sequence corresponding to a segment of the stalk region (amino acid, a.a., 336 to 590, Fig. 2a) of the motor protein kinesin²⁶. Although native kinesin is homodimeric, sedimentation equilibria experiments indicated that CC1 was predominantly tetrameric (data not shown). The circular dichroism (CD) spectrum of CC1 has the characteristic feature of an α -helical coiled coil (Fig. 2b). CC1 showed a major cooperative conformational change at 35 °C and a minor one at 65 °C (Fig. 2c). Stability experiments with the separate coiled-coil regions (our unpublished data) demonstrated that these two transitions were respectively due to the unfolding of the longer and the shorter coiled-coil regions of CC1 (Fig. 2a). The CD signal

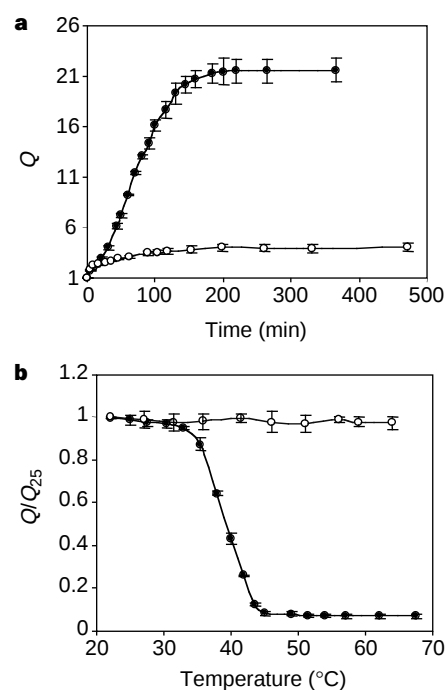


Figure 3 Dynamic swelling and the temperature-induced volume transition of the hybrid hydrogels in PBS. The extent of gel swelling was determined by measuring the two-dimensional changes of gel pieces of various shapes. For example, the changes of the three sides of a triangular gel piece were measured based on the optical images taken at different times during the course of gel swelling. The dimensional changes along different directions were found to be the same, within experimental error, indicating three-dimensional isotropic swelling of the hydrogel networks. Filled circles, gel 1; open circles: gel 2. **a**, Dynamic swelling characterized as the volume swelling ratio Q versus time. The one-dimensional swelling ratio is defined as L/L_0 , where L is the size of the swollen gel, and L_0 is the corresponding size of the dried gel. L and L_0 were determined from optical images of the gels. These images were obtained using a Nikon Eclipse E800 optical microscope, and photographs were taken with a CCD camera. Assuming three-dimensional isotropic swelling, the volume swelling ratio Q was calculated as $(L/L_0)^3$. **b**, Effect of temperature on the gel swelling behaviour. Q/Q_{25} is the ratio of the equilibrium volume swelling ratio of gels at an elevated temperature and the equilibrium volume swelling ratio at 25 °C in the same buffer. Gels were placed in a cuvette with a jacket connected to a water bath. The rate of temperature increase was 1 °C min⁻¹, with equilibration time of 2 min. The temperature of the water bath was calibrated by measuring the temperature inside the cuvette with a thermocouple. The error bars represent the standard deviation of three separate measurements.

of folded CC1 was fully restored after a temperature cycle, indicating a reversible folding–unfolding process.

The second protein, CC2, contains a *de novo* designed coiled-coil sequence (EK42) and a terminal His tag separated by a non-coiled-coil region of 30 amino acids (Fig. 2a). EK42 was previously designed to form parallel dimeric coiled coils²⁷, based on the principles that are thought to govern the folding and oligomerization of many native coiled coils and synthetic model peptides. Sedimentation equilibria measurements, however, revealed that CC2 forms tetramers in solution (data not shown). The CD spectrum of CC2 suggested an α -helical coiled-coil conformation (Fig. 2b). Compared with previous data on EK42²⁷, the thermal stability of CC2 was unexpectedly high, as shown by melting experiments (Fig. 2c). CC2 did melt, in a cooperative fashion similar to that observed for CC1, at temperatures under 90 °C in the presence of a denaturing agent, guanidine hydrochloride (GuHCl). The melting temperature (T_m) of CC2 was linearly related to the GuHCl concentration (data not shown). Thus we estimated the T_m of CC2 in the absence of GuHCl to be ~ 108 °C by extrapolation to zero GuHCl concentration. The unexpectedly high stability of CC2 may be due to the presence of the non-coiled-coil spacer sequence (Fig. 2a).

Hybrid hydrogels were assembled by mixing the chelating copolymer of HPMA charged with Ni^{2+} and the coiled-coil proteins, CC1 and CC2. Homogeneous films were obtained after drying the mixed solutions on Teflon sheets. Using the same procedure, control films were prepared with: (1) chelating copolymer and Ni^{2+} without coiled coils, (2) coiled coils without the chelating copolymer or Ni^{2+} , and (3) chelating copolymer and coiled coil without Ni^{2+} .

To verify the formation of hydrogels, the dried films were rehydrated in phosphate buffer (20 mM sodium phosphate, pH 7.4) in the absence of salt. The control films dissolved readily on immersion in the buffer. In contrast, the films formed with all the components—chelating HPMA copolymer, Ni^{2+} and the coiled coils—remained physically intact for 2 h before disintegrating due to excess swelling. Moreover, in the presence of 150 mM NaCl, these films remained physically intact for at least 24 h. Adding imidazole, which competes with the His tags for Ni^{2+} , increased the swelling of the gels at moderate concentrations (for example, 100 mM) and dissolved the gels at higher concentrations (such as 500 mM). These observations indicated that three-dimensional hybrid hydrogels were created as a result of the specific interactions proposed in Fig. 1. Gels assembled with CC1 and CC2 were designated as gel 1 and gel 2, respectively. Swelling profiles of these gels were studied in phosphate buffered saline (PBS; 20 mM sodium phosphate, 150 mM NaCl, pH 7.4) (Fig. 3a). The swelling was fast at first, more than doubling the dried gel volume in less than 10 min after being submerged in PBS. After about 3 h, swelling of the gels approached equilibrium. During the next 24 h, the gel volume increased slightly owing to the relaxation of the polymeric chains, as the force of hydration balanced the elastic constraint of the three-dimensional gel network. The same copolymer precursor was used in gel preparation and both coiled coils were similarly charged. The difference in swelling kinetics and equilibrium swelling ratio may be due to the considerable difference in the size of the proteins (Fig. 2a), the difference in gel crosslinking density, and/or the difference in the polymer–solvent interaction parameter, χ .

We investigated the temperature-responsiveness of the hybrid hydrogels containing CC1 and CC2, which exhibited different thermal stability in solution. On an increase in temperature from 25 °C to 70 °C, gel 1 underwent a sudden collapse to 10% of its equilibrium volume at 25 °C, with a mid-point transition temperature of 39 °C (Fig. 3b). This mid-point temperature of gel collapse was in good agreement (within 5 °C) with the T_m of the main coiled-coil region of CC1 as determined by CD. In contrast, no change in swelling was observed for gel 2 over 25 °C to 70 °C, as expected from the CD melting data for CC2 in solution.

We propose that the gel structural transition is caused by the temperature-induced conformational change in the CC1 coiled coil. CC1 is largely a rigid elongated ‘rod’ with extended coiled-coil structure; around 170 of the 255 amino acid residues of CC1 have strong propensity for adopting the coiled-coil conformation (Fig. 2a). The length of the CC1 bundle was estimated from the electron micrographs²⁸ to be about 25–30 nm. As the protein unfolds, there would be a considerable decrease in the hydrodynamic radius of the individual molecules from the initial elongated rods to random coils that caused the gels to collapse. Although such a conformational transition of CC1 was reversible in solution, the temperature-induced gel transition was not reversible within the relatively short period of time of the experiments (48 h). Physical constraints from the synthetic copolymer backbones and chain entanglement through exposed hydrophobic patches of the coiled coils could be contributing factors.

Our hybrid hydrogel system possesses several distinct features. The use of metal complexation as a means of anchoring His-tagged proteins should be applicable to all recombinant proteins with this affinity tag. Other metal ions, such as Zn^{2+} and Ga^{3+} , could be used instead of Ni^{2+} , for purposes such as stabilizing proteins, or tracking the hydrogels *in vivo* using magnetic resonance imaging (MRI). Unlike responsive gels formed solely by self-assembly of protein molecules containing α -helices²⁹ and β -sheets³⁰, this hybrid hydrogel system preserves the benefit of using synthetic polymer backbones that are well characterized, easy to manufacture and biocompatible. Furthermore, it offers the possibility of creating hydrogels of unusual physicochemical and/or biological properties by tailoring the coiled coils using techniques of molecular biology. For example, coiled coils with different T_m values could be used to assemble hybrid hydrogels that would display accordingly different gel structural transition temperatures. Moreover, two or more of these coiled coils could be used simultaneously, resulting in gels capable of stepwise transitions with each step triggered at a different temperature, which is difficult to achieve using hydrogels of synthetic polymers alone. The hybrid hydrogel system we describe here may also be useful for incorporating and delivering therapeutic proteins. Because these hydrogels are assembled in aqueous solvents, the biological activities of the protein drugs could be preserved. □

Methods

Construction of CC1 expression vector. A segment of DNA encoding the region of the *Drosophila* kinesin stalk from amino acid Val 336 to amino acid Ser 590 was PCR-amplified. To facilitate subcloning of the PCR product, the coding primer (5'-GGTCTAGAGTGGTCTGCGTTAACGAG-3') containing a terminal *Xba*I site and the non-coding primer (5'-CCCCGCGAGTCCA GCCTCGAGCC-3') containing a terminal *Ava*I site were used. Following digestion with *Xba*I and *Ava*I, the PCR product was ligated into the pET21a expression vector (Novagen, Madison, WI), which had been digested with *Nhe*I and *Ava*I. The CC1 vector expressed kinesin stalk protein which had, at the amino terminus, the additional sequence Met-Ala-Arg, and at the carboxy terminus the appended sequence Leu-Glu-His-His-His-His-His. The correct construct was verified by direct sequencing.

Construction of CC2 expression vector. An artificial gene encoding the EK42 domain was constructed from six synthesized single-stranded oligonucleotides by mixing the complementary strands to form three duplexes, which were ligated into the expression vector pRSETB (Invitrogen, Carlsbad, CA), cut with *Bam*HI and *Eco*RI, and modified to contain a kanamycin-resistance gene. The correct construct was verified by direct sequencing.

Protein expression and purification. Both proteins were expressed in *Escherichia coli* BL21(DE3)pLysS (Novagen, Madison, WI) using isopropyl β -thiogalactoside (IPTG) as the inducing agent, and were purified from soluble portion of cell lysate using Ni-NTA metal affinity resin (Qiagen, Santa Clarita, CA). SDS-polyacrylamide gel electrophoresis, amino-acid analysis and matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) were used to verify the identity and purity of the proteins.

Circular dichroism. The measurements were performed at 25 °C in PBS using an Aviv 62DS CD spectrometer equipped with a thermoelectric temperature control system and a 1.0-cm-pathlength quartz cuvette. Protein concentration used was 4 µM. Wavelength scans (1-nm bandwidth, 1 nm per step, and 5-s data averaging) for each sample were repeated five times, averaged, corrected by subtracting the buffer spectrum, and smoothed. For thermal melting experiments, the CD signal at 222 nm was monitored as temperature was increased from 10 to 90 °C with 0.1 or 2 °C per step. At each temperature, the sample was equilibrated for 5 min followed by 60 s of data point averaging. The transition temperatures were determined from the first-order differentiation of the melting curves.

Chemical syntheses. *N*-(*N'*,*N'*-dicarboxymethylaminopropyl)methacrylamide (DAMA) was synthesized by reacting bromoacetic acid (14 mmol, 2.0 g) with *N*-(3-aminopropyl)methacrylamide hydrochloride (7 mmol, 1.25 g) at alkaline pH. The reaction was allowed to progress for 2 h at 0 °C, then for 48 h at room temperature. DAMA was recrystallized from anhydrous ethanol; yield, 0.5 g (30%); melting point, 178–180 °C. DAMA was characterized by thin layer chromatography, mass spectrometry, proton nuclear magnetic resonance and elemental analysis. Copolymerization of DAMA and HPMA was performed in methanol (86.9 wt%) with 2,2'-azobisisobutyronitrile (AIBN, 0.6 wt%) as initiator. The reaction progressed in nitrogen atmosphere at 50 °C for 24 h. The copolymer was precipitated into acetone, dialysed against water, and lyophilized. Relative molecular mass (240,000) and polydispersity (1.68) were measured by size exclusion chromatography on a fast protein liquid chromatography system equipped with a Superose 6 (10/30) column (Pharmacia, Piscataway, NJ) and a light-scattering detector (MiniDawn, Wyatt). The column was calibrated using polyHPMA fractions with small polydispersity. The content of carboxylic groups (15 mol%) was determined by acid–base titration using a 10-ml ABU80 autoburette and a PHM84 research pH meter.

Hybrid hydrogel preparation. To prepare the metal-charged copolymer precursor, a solution of poly(HPMA-co-DAMA) (200 mg dissolved in 10 ml of deionized water with pH adjusted by 1 M NaOH to 7.5–7.8) was mixed with excess Ni²⁺ (10 ml of 0.1 M NiSO₄ with pH adjusted by NaOH to 7.1) under constant stirring, in order to prevent premature crosslinking of the copolymer precursors. The copolymer–Ni²⁺ complex was separated from the excess free Ni²⁺ using a size exclusion column packed with Sephadex G-25 beads (Pharmacia, Piscataway, NJ) equilibrated with deionized water. To prepare the hybrid hydrogels, the coiled-coil proteins (2 mg of each dissolved in 0.1 ml of deionized water) were separately added to, and mixed with, solutions of the copolymer–Ni²⁺ complex (0.55 mg in 0.1 ml of deionized water for gel 1, and 1.1 mg in 0.2 ml of deionized water for gel 2). The mixed solutions were dropped on a Teflon sheet with 50 µl per drop, and dried at room temperature and atmospheric pressure. Thin films of gels formed were circular, with diameter of 5 mm and thickness of ~0.1 mm. They were cut and stored in a desiccator at 25 °C for use in swelling studies.

Received 14 September; accepted 27 November 1998.

1. Langer, R. Drug delivery and targeting. *Nature* **392** (suppl.), 5–10 (1998).
2. Peppas, N. A. Hydrogels and drug delivery. *Curr. Opin. Colloid Interface Sci.* **2**, 531–537 (1997).
3. Kopeček, J., Vacík, J. & Lim, D. Permeability of membranes containing ionogenic group. *J. Polym. Sci. A-1*, **9**, 2801–2815 (1971).
4. Suzuki, A. & Tanaka, T. Phase transition in polymer gels induced by visible light. *Nature* **346**, 345–347 (1990).
5. Kokufuta, E., Zhang, Y. Q. & Tanaka, T. Saccharide-sensitive phase transition of a lectin-loaded gel. *Nature* **351**, 302–304 (1991).
6. Kwon, I. C., Bae, Y. H. & Kim, S. W. Electrically erodible polymer gel for controlled release of drugs. *Nature* **354**, 291–293 (1991).
7. Osada, Y., Okuzaki, H. & Hori, H. A polymer gel with electrically driven motility. *Nature* **355**, 242–244 (1992).
8. Chen, G. & Hoffman, A. S. Graft copolymers that exhibit temperature-induced phase transitions over a wide range of pH. *Nature* **373**, 49–52 (1995).
9. Yoshida, R. *et al.* Comb-type grafted hydrogels with rapid de-swelling response to temperature changes. *Nature* **374**, 240–242 (1995).
10. Holtz, J. H. & Asher, S. A. Polymerized colloidal crystal hydrogel films as intelligent chemical sensing materials. *Nature* **389**, 829–832 (1997).
11. Hu, Z., Chen, Y., Wang, C., Zheng, Y. & Li, Y. Polymer gels with engineered environmentally responsive surface patterns. *Nature* **393**, 149–152 (1998).
12. Kiser, P. F., Wilson, G. & Needham, D. A synthetic mimic of the secretory granule for drug delivery. *Nature* **394**, 459–462 (1998).
13. Šubr, V., Duncan, R. & Kopeček, J. Release of macromolecules and daunomycin from hydrophilic gels containing enzymatically degradable bonds. *J. Biomater. Sci. Polym. Edn* **1**, 261–278 (1990).
14. Nagahara, S. & Matsuda, T. Hydrogel formation via hybridization of oligonucleotides derivatized in water-soluble vinyl polymers. *Polym. Gels Networks* **4**, 111–127 (1996).
15. Obaidat, A. A. & Park, K. Characterization of glucose dependent gel-sol phase transition of the polymeric glucose-concanavalin A hydrogel system. *Pharm. Res.* **13**, 989–995 (1996).

16. Wang, C., Stewart, R. J. & Kopeček, J. De novo design of hybrid hydrogels: water soluble polymers crosslinked by coiled-coil protein domains. *Proc. Int. Symp. Controlled Release Bioact. Mater.* **25**, 54–55 (1998).
17. Wang, C., Stewart, R. J. & Kopeček, J. Tailor-made hybrid hydrogels: synthetic macromolecules crosslinked by coiled-coil protein domains. *ACS Polym. Preprints* **39**, 194–195 (1998).
18. Lupas, A. Coiled coils: new structures and new functions. *Trends Biochem. Sci.* **21**, 375–382 (1996).
19. O'Shea, E. K., Rutkowski, R. & Kim, P. S. Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. *Cell* **68**, 699–708 (1992).
20. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **262**, 1401–1407 (1993).
21. Su, J. Y., Hodges, R. S. & Kay, C. M. Effect of chain length on the formation and stability of synthetic α -helical coiled coils. *Biochemistry* **33**, 15501–15510 (1994).
22. Gonzalez, L. Jr, Plecs, J. J. & Alber, T. An engineered allosteric switch in leucine-zipper oligomerization. *Nature Struct. Biol.* **3**, 510–515 (1996).
23. Hochuli, H. Purification of recombinant proteins with metal chelate adsorbent. *Genet. Eng.* **12**, 87–98 (1990).
24. Ng, K., Pack, D. W., Sasaki, D. Y. & Arnold, F. H. Engineered protein-lipid interactions: targeting of histidine-tagged proteins to metal-chelating lipid monolayers. *Langmuir* **11**, 4048–4055 (1995).
25. Ho, C. H., Limberis, L., Caldwell, K. D. & Stewart, R. J. A metal-chelating pluronic for immobilization of histidine-tagged proteins at interfaces: immobilization of firefly luciferase on polystyrene beads. *Langmuir* **14**, 3889–3894 (1998).
26. Yang, J. T., Laymon, R. A. & Goldstein, L. S. B. A three-domain structure of kinesin heavy chain revealed by DNA sequence and microtubule binding analyses. *Cell* **56**, 879–889 (1989).
27. Graddis, T. J., Myszkowski, D. G. & Chaiken, I. M. Controlled formation of model homo- and heterodimer coiled coil polypeptides. *Biochemistry* **32**, 12664–12671 (1993).
28. Hirokawa, N. *et al.* Submolecular domains of bovine brain kinesin identified by electron microscopy and monoclonal antibody decoration. *Cell* **56**, 867–878 (1989).
29. Petka, W. A., Harden, J. L., McGrath, K. P., Wirtz, D. & Tirrell, D. A. Reversible hydrogels from self-assembling artificial proteins. *Science* **281**, 389–392 (1998).
30. Aggeli, A. *et al.* Responsive gels formed by the spontaneous self-assembly of peptides into polymeric β -sheet tapes. *Nature* **386**, 259–262 (1997).

Acknowledgements. This work was supported in part by the Center for Biopolymers at Interfaces (CBI) and the University of Utah Research Foundation.

Correspondence and requests for materials should be addressed to J.K. (e-mail: Jindrich.Kopecek@mcc.utah.edu).

A complex clathrate hydrate structure showing bimodal guest hydration

Konstantin A. Udachin & John A. Ripmeester

Steacie Institute for Molecular Sciences, National Research Council of Canada, Ottawa, Ontario, Canada K1A 0R6

Interactions between hydrophobic groups in water¹, as well as biomolecular hydration more generally^{2–4}, are intimately connected to the structure of liquid water around hydrophobic solutes. Such considerations have focused interest on clathrate hydrates: crystals in which a hydrogen-bonded network of water molecules encages hydrophobic guest molecules with which the water interacts only by non-directional van der Waals forces. Three structural families of clathrate hydrates have hitherto been recognized: cubic structure I (2M_S·6M_L·46H₂O) (ref. 5), cubic structure II (16M_S·8M_L·136H₂O) (ref. 5) and hexagonal structure H (M_L·3M_S·2M_S·34H₂O) (refs 6, 7) hydrates (here M_L and M_S are the hydrophobic guest sites associated with large and small cavities, respectively). Here we report a new hydrate structure: 1.67 choline hydroxide-tetra-*n*-propylammonium fluoride·30.33H₂O. This structure has a number of unusual features; in particular the choline guest exhibits both hydrophobic and hydrophilic modes of hydration. Formally the structure consists of alternating stacks of structure H and structure II hydrates, and might conceivably be found in those settings (such as seafloor deposits over natural-gas fields) in which clathrate hydrates form naturally.

The classical clathrate structures are three-dimensional lattices that can be built up by fitting together different types of closed cages composed of 4-, 5- and 6-sided polygons (Table 1). Both in experimental work and modelling studies on biomolecular hydration, structural features such as closed pentagonal rings of hydrogen-bonded water molecules are taken as evidence for clathrate-like hydrophobic hydration^{2–4}. Although this notion generally

Table 1 Properties of hydrate structure types

Structure type	Unit-cell content Space group	Cage	Symmetry of cages	Number* of square faces	Number* of pentagonal faces	Number* of hexagonal faces
Structure I	2D·6T·46H ₂ O <i>Pm3n</i>	5 ¹² (D) 5 ¹² 6 ² (T)	m ³ 42m	0	0.96	0.13
Structure II	16D·8H·136H ₂ O <i>Fd3m</i>	5 ¹² (D) 5 ¹² 6 ⁴ (H)	3m 43m	0	1.06	0.12
Structure H	E·3D·2D'·34H ₂ O <i>P6/mmm</i>	5 ¹² (D) 4 ³ 5 ⁶ 6 ³ (D') 5 ¹² 6 ⁸ (E)	mmm 6m2 6/mmm	0.09	0.88	0.21
New structure (ideal)	3E·12H·6D'·33D·306H ₂ O <i>R-3m</i>	5 ¹² (D) 4 ³ 5 ⁶ 6 ³ (D') 5 ¹² 6 ⁴ (H) 5 ¹² 6 ⁸ (E)	3, 3m 3m 3m, 3m 3m	0.03	1.0	0.15
New structure (actual)	3E·12H·3D·3D ₄ ·6D ₃ D'·306H ₂ O <i>R-3</i>	5 ¹² (D) 4 ³ 5 ³⁰ 6 ³ (D ₄) 5 ³⁶ (D ₃ D') 5 ¹² 6 ⁴ (H) 5 ¹² 6 ⁸ (E)	1, 3 3 3 3, 3 3	0.03	0.82	0.15

* Normalized by the number of water molecules per unit cell.

is correct, Table 1 shows that, with the incorporation of even completely hydrophobic species, the number of polygons different from pentagons increases as the guest size increases (structures I, II → structure H).

This work was undertaken as a first step to building a bridge between these classical hydration structures, and structures expected to display biomolecular hydration. Rather than examining hydration structures associated with proteins or peptides, we take a simple biomolecular fragment and incorporate it into a clathrate lattice. We have chosen choline hydroxide, choline ((CH₃)₃N⁺CH₂CH₂OH) being an important fragment in a variety of complex phospholipids. When choline hydroxide was co-crystallized with tetra-*n*-propylammonium fluoride from aqueous solution, an unusual complex hydrate structure resulted. The second component was chosen as a 'help-molecule' because it is not known to form a stable clathrate hydrate⁸, yet it is closely related to the tetra-*n*-alkylammonium salts that form a variety of hydrate structures⁹. One unit cell is shown in Fig. 1. The *c*-dimension of ~90 Å is remarkable, and we will show that this structure in fact consists of sequences of layers characteristic of other hydrate structures stacked along the *c* direction. Further details of the

crystallographic study are given in Methods, in Table 2 and in Supplementary Information.

The tetra-*n*-propylammonium guest is located in clusters of D' and D cages fused to form large D₄ and D₃D' supercages (Fig. 2a), an arrangement observed for other *n*-alkylammonium environments in salt hydrates⁸. In these two crystallographically distinct supercages, each tetra-*n*-propylammonium guest has three different possible orientations of the alkyl chain, although only one is shown in the figure. Choline is located in both of the large cage types, E and H, present in the lattice (see Fig. 1), with cage occupancies of 1 and 0.5, respectively. The H cage occurs as two crystallographically distinct versions; again, the guest is disordered over several positions, and we show only a single orientation (Fig. 2b). Two competing processes are expected to play a role in locating the choline guest molecule inside the cavity. On the one hand, the molecule is expected to arrange itself in the cavity so as to maximize non-specific van der Waals contacts with the cavity walls. On the other hand, the hydroxyl group of the guest molecule can be expected to form a strongly directional hydrogen bond with the water framework. The H cavity is too small to accommodate the guest molecule completely in a hydrophobic way, and the guest hydroxyl displaces one of the host-lattice water molecules. But owing to its tendency to maximize van der Waals contacts with the cage, the choline molecule is arranged in one H cavity in such a way that the hydroxyl group is 1.3 Å from the ideal location of the displaced water molecule. Thus, the choline molecule lies in an

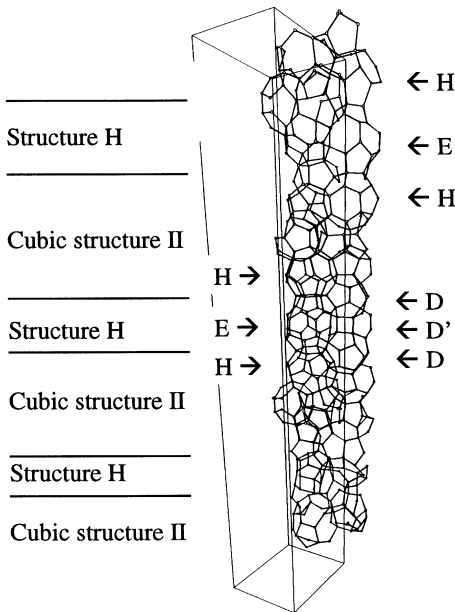


Figure 1 Unit cell of 1.67choline hydroxide-tetra-*n*-propylammonium fluoride-30.33H₂O. Cage types and stacking sequences are indicated (see text).

Table 2 Crystal data and structure refinement summary

Empirical formula	C ₁₆ H _{99.66} FN ₂ O _{32.33}
Formula weight	856.92
Temperature	243(2) K
Wavelength	0.71073 Å
Crystal system, space group	Trigonal, <i>R-3</i>
Unit cell dimensions	<i>a</i> = 12.53350(10) Å, <i>α</i> = 90° <i>b</i> = 12.53350(10) Å, <i>β</i> = 90° <i>c</i> = 90.5248(12) Å, <i>γ</i> = 120°
Volume	12,315.2(2) Å ³
<i>Z</i>	12
Calculated density	1.387 Mg m ⁻³
Crystal size	0.40 × 0.40 × 0.20 mm
Reflections collected/unique	28,033/6,972 (<i>R</i> (int) = 0.0498)
Goodness-of-fit on <i>F</i> ²	1.011
Final <i>R</i> indices (<i>I</i> > 2σ(<i>I</i>))	<i>R</i> 1 = 0.1092, <i>wR</i> 2 = 0.2997

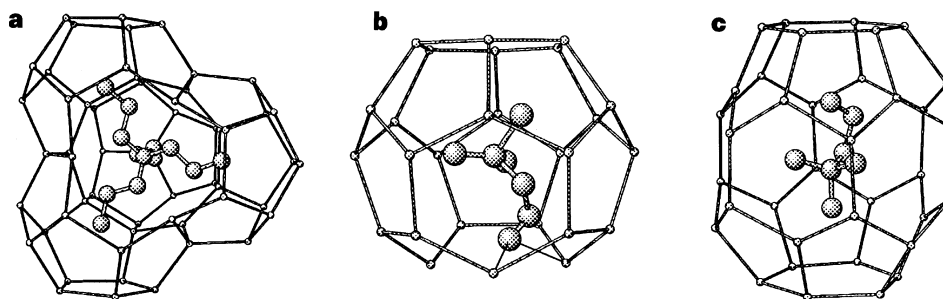


Figure 2 Guest molecules in the cages of the hydrate structure. **a**, Tetra-*n*-propylammonium ion in the D₃D' fused supercage. The guest is shown in one of three disordered orientations. **b**, Choline fragment in one of the two crystallographically distinct H cages. The hydroxyl oxygen is shown replacing one of the

cage water molecules, forming hydrogen bonds to two water molecules. **c**, Choline fragment in the E cage. There are only van der Waals contacts between the guest and the host cage.

intermediate position, forming the maximum number of van der Waals contacts, and also forming hydrogen bonds with O–O distances of 2.84 Å and with OHO angles of 75°, 96° and 134° (Fig. 2b). In the other H cavity, the hydroxyl group is displaced by 1.9 Å, and the corresponding hydrogen-bond distances and angles are 2.80 Å, 78°, 94° and 146°.

The choline guest in the other large cage of the structure, the E cage (see Fig. 1), is also present in a disordered fashion but it does not interact with the cage by hydrogen bonding at all, there being only van der Waals contacts (Fig. 2c): the two shortest distances from the hydroxyl oxygen to cage water molecules are 3.38 Å (O1K–O15) and 3.75 Å (O1K–O14) (see Supplementary Information). The structure also contains empty D cages, which are analogous to those found in structure H hydrate. The fluoride and hydroxyl anions, necessary for charge balance, cannot be located, as the diffraction experiment does not distinguish them from water molecules; this is much as observed for other hydrates containing these anions.

One conclusion suggested by these results is that the nature of hydration experienced either by small molecules or by components of larger molecules cannot be predicted in a simple way. One would assume hydrogen bonding to be more effective than non-directional weak guest–host interactions as the mode of hydration of fragments containing hydroxyl functions, yet it must be the minimum lattice energy attained by the entire structure that dictates whether any particular fragment will follow a particular mode. In this context, we now consider current opinion on hydrophobic hydration. It has been recognized for some time that in solution, many-body interactions give rise to long-range correlated solvent-induced forces; this means that individual pair-wise interactions no longer have much meaning¹⁰. This would seem to affirm that two very different hydration modes are able to co-exist in solution as well as in the solid state, as pair-wise interactions would suggest the predominance of a single mode of hydration. Of course, it is possible that the (as yet unknown) fluoride and hydroxide locations also play a role in deciding on a particular hydration mode for the different choline fragments.

A number of hydrate structures, including structure II, structure H and structure IV in Jeffrey's classification⁵, can formally be constructed from the stacking of two-dimensional sheets consisting of 5¹² cages joined by sharing a pentagonal face. Structure H and structure IV hydrates are obtained by AA stacking, and cubic structure II hydrate by ABCABC stacking, of such sheets. A hypothetical hexagonal version of structure II hydrate can be constructed by ABAB stacking. On closer scrutiny, the present hydrate structure also can be analysed in terms of stacked sequences of layers, and the scheme that emerges for this structure is CABBCAABCCABBCAABC. Such a sequence is unprecedented, either for hydrate structures¹¹ or the related silicate lattices known

as clathrasils¹². Evidence for all three clathrate structures, mainly with hydrocarbon guests^{13,14}, has been found in natural settings; in this connection, these materials have attracted a great deal of attention as potential energy sources¹⁵, as possible agents for global climate change¹⁵, and as hazards in the pipeline transport of hydrocarbons¹⁶. The closed-cage version of this structure, E·4H·2D'·11D·102H₂O, where guests are not hydrogen-bonded to the cages and the D and D' cages are complete, is another structure that possibly may be found in natural or industrial settings. The range of cavity sizes present would allow the versatile incorporation of a wide variety of naturally occurring molecules: methane, H₂S, O₂, N₂, CO₂ in the D and D' cages, ethane, propane, isobutane and butane in the H cage, and larger hydrocarbons such as neohexane and isopentane in the E cage. The existence of a stable structure of this kind would depend on its stability with respect to the stability of the individual simpler hydrate structures. Other possibilities that should be considered include intergrowths of structure II, H and the new structure, as these are all based on the stacking of layers of pentagonal dodecahedra.

The data in Table 1 also show why hydration structures containing large molecules can be expected to have fewer pentagonal rings than the ideal clathrates², even though the structures are closely related. The large cages (D₄, D₃D') formed from fused smaller cages quite naturally have fewer faces of this kind, as the internal walls formed by the pentagonal faces are missing. The new hydrate structure reported here, because of the relatively small units incorporated, can still maintain a regular periodic structure that would be far more difficult to attain for irregularly shaped large biomolecules. □

Methods

1.67Choline hydroxide-tetra-*n*-propylammonium fluoride·30.33H₂O was crystallized from aqueous solution at –3 °C. A crystal with the morphology of a hexagonal prism (plate) of dimensions 0.2 × 0.6 × 0.6 mm was chosen for the diffraction experiment. Single-crystal diffraction data were collected on a Siemens SMART CCD diffractometer (temperature 243(2) K), and the structure solved using the SHELXTL package^{17,18}. For further details see Table 2; bond distances and angles, fractional coordinates and anisotropic displacement parameters are available as Supplementary Information.

Received 22 June; accepted 2 November 1998.

1. Blokzy, W. & Engberts, J. B. F. Hydrophobic effects. Opinions and facts. *Angew. Chem. Int. Edn Engl.* **32**, 1545–1579 (1993).
2. Head-Gordon, T. Is water structure around hydrophobic groups clathrate-like? *Proc. Natl Acad. Sci. USA* **92**, 8308–8312 (1995).
3. Lipscomb, L. A., Zhou, F. X. & Williams, L. D. Clathrate hydrates are poor models of biomolecule hydration. *Biopolymers* **38**, 177–181 (1996).
4. Bouquiere, J. P., Finney, J. L. & Savage, H. F. J. High-resolution neutron study of vitamin B12 coenzyme at 15K: solvent structure. *Acta Crystallogr. B* **50**, 566–578 (1994).
5. Jeffrey, G. A. in *Inclusion Compounds* Vol. 1 (eds Atwood, J. L., Davies, J. E. D. & MacNicol, D. D.) 135–190 (Academic, London, 1984).
6. Ripmeester, J. A., Ratcliffe, C. I., Tse, J. S. & Powell, B. M. A new clathrate hydrate structure. *Nature* **325**, 135–136 (1987).

7. Udachin, K. A., Ratcliffe, C. I., Enright, G. D. & Ripmeester, J. A. Structure H hydrate: a single crystal diffraction study of 2,2 dimethylpentane-5(Xe,H₂S).34H₂O. *Supramol. Chem.* **8**, 173–176 (1997).
8. Lipkowsky, J., Luboradzki, R., Udachin, K. A. & Dyadin, Yu. A. A layered hydrate structure of tetrapropyl ammonium fluoride. *J. Inclusion Phenom.* **13**, 295 (1992).
9. Dyadin, Yu. A., Udachin, K. A., Zhurko, F. V. & Mironov, Yu. I. Cubic structure II double clathrate hydrates with tetra-*n*-propylammonium fluoride. *J. Inclusion Phenom.* **6**, 565–575 (1988).
10. Martorana, V., Bulone, D., San Biagio, P. L., Palma-Vitorelli, M. B. & Palma, M. U. Collective properties of hydration—long range and specificity of hydrophobic interactions. *Biophys. J.* **73**, 31–37 (1997).
11. Smelik, E. A. & King, H. E. Jr Clathrate hydrates as layer structures. Comparison between the structures of type II and hypothetical type V. *Z. Kristallogr.* **211**, 84–89 (1996).
12. Gies, H. in *Inclusion Compounds* Vol. 5 (eds Atwood, J. L., Davies, J. E. D. & MacNicol, D. D.) 1–31 (Oxford Univ. Press, 1991).
13. Davidson, D. W. *et al.* Laboratory analysis of a naturally occurring gas hydrate. *Geochim. Cosmochim. Acta* **50**, 619–623 (1986).
14. Sassen, R. & MacDonald, I. R. Evidence of structure H hydrate. Gulf of Mexico continental shelf. *Org. Geochem.* **22**, 1029–1032 (1994).
15. Kvenvolden, K. A. Natural gas hydrate occurrence and issues. *Ann. N.Y. Acad. Sci.* **715**, 232–247 (1994).
16. Sloan, E. D. Jr *Clathrate Hydrates of Natural Gas* 2nd edn (Dekker, New York, 1996).
17. Sheldrick, G. M. Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallogr. A* **46**, 467–473 (1990).
18. Sheldrick, G. M. A new-least squares refinement program for use with single crystal diffraction data. *Acta Crystallogr. A* **49** (Suppl.), C53 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank the NATO Science Programme for partial support of this work in the form of a Collaborative Research Grant to K.A.U.

Correspondence and requests for materials should be addressed to J.A.R. (e-mail: jar@ned1.sims.nrc.ca).

Upward transport of oceanic nitrate by migrating diatom mats

Tracy A. Villareal*, Cynthia Pilskaln†, Mark Brzezinski‡, Fredric Lipschultz§, Mark Dennett|| & George B. Gardner¶

* Marine Science Institute, The University of Texas at Austin, 750 Channel View Drive, Port Aransas, Texas 78373, USA

† School of Marine Sciences, 5741 Libby Hall, University of Maine, Orono, Maine 04469, USA

‡ Marine Science Institute and Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, California 93106, USA

§ Bermuda Biological Station for Research, Inc., 17 Biological Station Lane, Ferry Reach, GE01 Bermuda

|| Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

¶ Environmental, Coastal and Ocean Sciences Program, University of Massachusetts, 100 Morrissey Boulevard, Boston, Massachusetts 02125, USA

The oligotrophic gyres of the open sea are home to a flora that includes the largest known phytoplankton. These rare species migrate as solitary cells or aggregations (mats) between deep nutrient pools (below 80–100 m) and the surface. This migration contributes to new production because of the concomitant upward transport of nitrate^{1–3}. But just how significant this contribution is remains uncertain because of the difficulty of making quantitative measurements of these rare cells⁴. Here we report remote video observations of a previously undersampled class of diatom (*Rhizosolenia*) mats throughout the upper 150 m of the central North Pacific Ocean. These mats are virtually invisible to divers, and their presence increases the calculated phytoplankton-mediated nitrate transport into the surface ocean by up to a factor of eight. Cruise averages indicate that *Rhizosolenia* mats transport 18–97 $\mu\text{mol N m}^{-2} \text{ d}^{-1}$; however, this value reached 171 $\mu\text{mol N m}^{-2} \text{ d}^{-1}$ at individual stations, a value equivalent to 59% of the export production⁵. Although considerable temporal and spatial variability occurs, this means of upward nutrient transport appears to be an important source of new nitrogen to the surface ocean, and may contribute to other regional elemental cycles as well⁶.

Rhizosolenia mats (macroscopic associations of *Rhizosolenia* spp. up to 30 cm in size^{7,8}) are representative of a group of non-motile oceanic diatom, dinoflagellate and prasinophyte taxa that have a vertical migration life history^{1,2,8}. Their transit to depth for nutrient acquisition and their return to the surface for photosynthesis has been inferred from a variety of isotopic, biochemical and behavioural data^{2,3}; however, no direct evidence confirms their existence in the depth range of the nitracline (the zone where nitrate concentration increases from nanomolar to micromolar concentrations). Their buoyancy-regulated migration has important biogeochemical implications as nitrate (new nitrogen) is transported from depth into the surface waters² and surface-derived carbon is respired at depth during migration⁴. This process circumvents the nutrient trap at the base of the euphotic zone⁹ and augments the flux of new nitrogen into, and carbon out of, the surface layer. In contrast, upwelled water transports both carbon and inorganic nitrogen upwards. Vertical migration is also one of several possible explanations for a preformed nitrate layer found just below the euphotic zone in the eastern basin of the gyre¹⁰. This layer is formed when nitrate uptake occurs with respiration yielding anomalous relationships between oxygen and nitrate concentrations. Field and model estimates suggested that vertical transport by *Rhizosolenia* mats in the central North Pacific gyre is only 17–25% of that due to turbulent nitrate diffusion⁴, but abundance estimates crucial for transport calculations have been severely constrained by depth limitations for SCUBA divers (20 m).

We studied *Rhizosolenia* mats using a video plankton recorder (VPR) at three stations in the central North Pacific gyre during June, 1996. These stations were typical of this gyre in the summer, with a nutricline present at 80–140 m depth (Fig. 1). Our hydrographic and analytical protocols can be found elsewhere^{2,3,6,11}. The VPR recorded clearly identifiable mats to 150 m depth, but of a size class rarely seen before (Fig. 2). Mats found by divers were 3–4 cm or greater in size; divers rarely noted mats <1 cm in size at the surface, and never saw small mats at >10 m depth. Only 25% of the VPR mats were >2.5 cm in size, and 60% were <1.5 cm in size.

Vertical profiles indicated a surface maximum in VPR mat abundance (Fig. 2a), but with abundance 10–20 times greater than diver estimates (Fig. 2b). Integrated values from divers were only 1.4–7.7 mats m^{-2} over the upper 20 m of the ocean, compared to 31–246 mats m^{-2} found by the VPR (Table 1). Integrated VPR mat abundance (0–150 m depth) yielded 80–999 mats m^{-2} with a maximum at station 3 (Table 1).

Small mats have been reported previously, but they are extremely difficult to see with the unaided eye due to their poor contrast⁸. The VPR's intense lighting and optical magnification highlighted the mats, and detailed their characteristic morphology (Fig. 2c). It is unlikely that the large mats noted by divers were even observed by the VPR, as this instrument examined only 4,011–5,315 litres, while divers sampled more than 50,000–80,000 litres per station. The high abundance of mats that we report here is not an artefact but a quantification of a poorly documented size class of diatom mats in the open sea.

We have revised previous estimates of the upward N transport using the VPR abundance data. Mat N content was scaled assuming an equivalent spherical volume, such that that a 1.25-cm mat will have 7.4% of the N content of a 3-cm mat³. This value may be as high as 16%, or 2.2 times our estimated value based on length–volume relationships from archived photographs¹². Our conservative comparisons indicate that the small mats represent 1.7–9.5 (average 4.5) times the N present in the larger mats over the depth range (0–20 m) where the measurements overlap (Table 1), and from 15 to 164 $\mu\text{mol N m}^{-2}$ in a 150-m water column. Using a biomass-specific N transport of 0.14 d^{-1} (ref. 4), the VPR abundance data (0–150 m) yielded an absolute transport rate of 2–23 $\mu\text{mol N m}^{-2} \text{ d}^{-1}$, or 1–8 (average, 4) times the transport based on diver (0–20 m) estimates (Table 1). The summed transport of

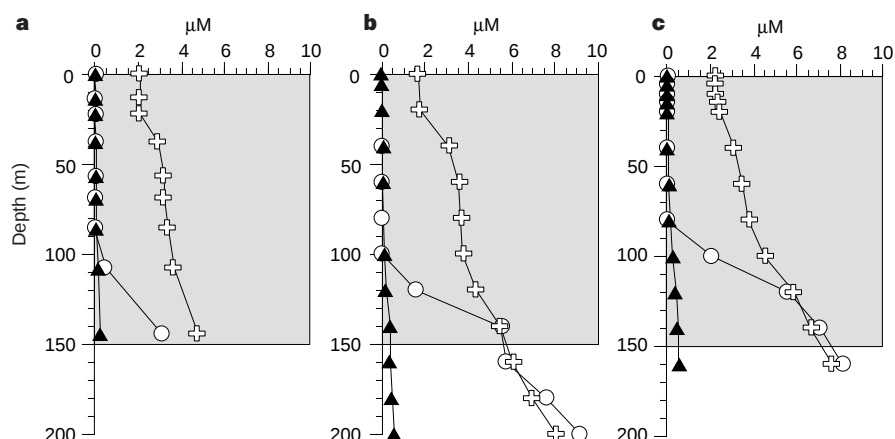


Figure 1 Nutrient profiles from the VPR stations. **a**, Station 3; **b**, station 4; **c**, station 5. Crosses, silicate; open circles, nitrate + nitrite; filled triangles, orthophosphate. The shaded area represents the depth range sampled by the VPR.

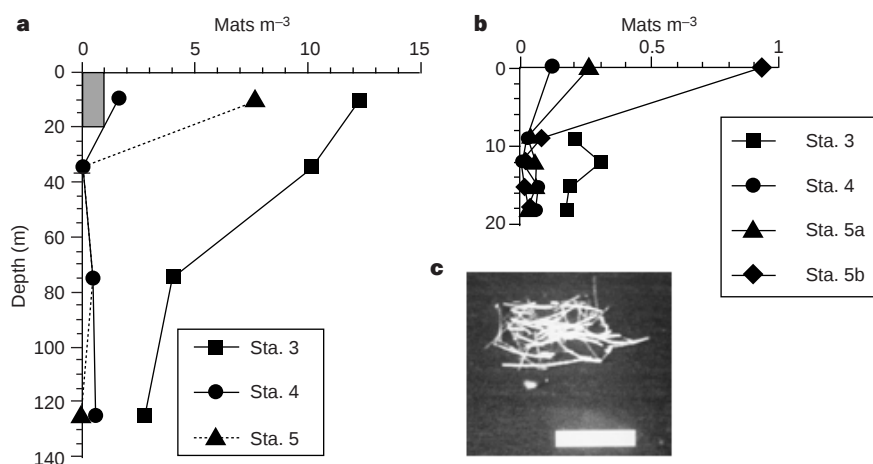


Figure 2 *Rhizosolenia* mat abundance estimates based on VPR measurements and reports from divers. **a**, VPR estimates. The shaded area represents the abundance and depth range of the abundance estimates from divers, reported in **b**. We note that in **b**, there is a different scale on both the x and y axes. Stations 5a and 5b represent respectively dives taken before and after the VPR sampling at station 5, and within 10 km of the VPR sampling site. **c**, Digitized video imagery of *Rhizosolenia* mats collected at 54 m depth (scale bar, 1 cm).

large mats (observed by divers) and small mats (measured by VPR) ranges from 4 to 26 $\mu\text{mol N m}^{-2} \text{d}^{-1}$.

In Table 2 we present multi-year N transport estimates for the eastern North Pacific central gyre using these new data (Table 2). Considerable temporal and spatial variability occurs, and ranges from 18 to 97 $\mu\text{mol N m}^{-2} \text{d}^{-1}$ over the 16-year period with an average of 40 $\mu\text{mol N m}^{-2} \text{d}^{-1}$. The individual stations show considerable intra-cruise variability, ranging from 0 to 171 $\mu\text{mol N m}^{-2} \text{d}^{-1}$. For comparison, turbulent nitrate fluxes estimates are in the range 22–1,644 $\mu\text{mol N m}^{-2} \text{d}^{-1}$, with 200 $\mu\text{mol N m}^{-2} \text{d}^{-1}$ suggested as being representative⁴. Average values suggest that migrating diatom mats transport N equal to 9–48% of the turbulent diffusive flux, or, on average, 20% of the representative value. However, at individual stations on three of the six cruises, the biological transport exceeded 75% of the average turbulent diffusive flux. An additional comparison can be made using the Hawaii Ocean Time-series (HOTS) data from 22° 45' N, 158° W (ref. 5). The 7-year particulate export production (losses from the euphotic zone) measurements, from the HOTS data, averaged 290 $\mu\text{mol N m}^{-2} \text{d}^{-1}$. *Rhizosolenia* mat N transport ranges from 6 to 33% of this flux at a site considered representative of the open Pacific Ocean^{13,14}. The maximum transport value equalled 59% of the HOTS export production, with average transport being 14% of export production. Using HOTS particulate C-fixation rates¹⁵, we calculate that average mat NO_3^- input equals 32% of

the mixed layer (0–40 m) new nitrogen needs, assuming an “*f*” ratio of 0.05 (see refs 16, 17). Such additional inputs, possibly by either N transport or N_2 fixation⁵, are required to support observed new production rates¹⁸, and underline the importance of biological introduction of nutrients into the euphotic zone.

Our calculations do not include the other solitary (non-aggregated) taxa known to migrate¹, but recent observations at frontal zones suggest that these other taxa may also occur at high concentrations in the open sea¹⁹. *Rhizosolenia* mat distribution in the North Pacific extends at least from 130° to 173° W and 20°–31° N, a range centred in the negative preformed NO_3^- zone, suggesting that these mats may be important in the development of this feature. *Rhizosolenia* mats are poorly documented outside the eastern North Pacific gyre, and we cannot extrapolate our results to other basins.

The wide range in transport estimates is due partly to analytical and conceptual uncertainties, but also represents true environmental variability. Transient mesoscale eddies contribute substantially to new N fluxes in the Sargasso Sea²⁰, and similar processes may occur in the North Pacific²¹. Direct comparison of biological transport and physical processes based on average values of either process may be of limited usefulness. Despite the uncertainties, these calculations emphasize the ability of large, rare cells to contribute significantly to new production, and also show the value of VPR technology in understanding the temporal and spatial importance of large, rare phytoplankton to oceanic biogeochemistry. □

Table 1 Comparison of diatom mat abundance estimates

Biomass distribution	Sta. 3	Sta. 4	Sta. 5	Average
VPR (0–20 m) mats m ⁻²	246	31	204	160.4
Diver (0–20 m) mats m ⁻²	7.7	1.4	1.6	3.5
VPR (0–150 m) mats m ⁻²	888	80	231	400.0
VPR (0–150 m) $\mu\text{mol N m}^{-2}$	164	15	43	74.0
Ratio VPR/diver mat N (0–20 m)	2.4	1.7	9.5	4.5
Transport calculations ($\mu\text{mol N m}^{-2} \text{d}^{-1}$)				
Diver (0–20 m)	3	2	2	2
VPR (0–150 m)	23	2	6	10
Ratio VPR/diver	8	1	3	4
Average transport of total diver (0–20 m) + VPR mats (0–150 m)	26	4	8	12

Biomass distribution in the VPR- and diver-measured fraction was calculated using an N content of 0.19 and 2.5 $\mu\text{mol N per mat}^{-1}$ for the former and latter mats, respectively. All data are for August 1996.

Table 2 Estimated upward new N transport due to migrating *Rhizosolenia* mats

Date	Minimum ($\mu\text{mol N m}^{-2} \text{d}^{-1}$)	Maximum ($\mu\text{mol N m}^{-2} \text{d}^{-1}$)	Average ($\pm$ s.d.) ($\mu\text{mol N m}^{-2} \text{d}^{-1}$)
Oct. 1980 ²³	40	160	97 $\pm$ 29
Sept. 1981 ²⁵	—	—	20
Apr. 1989 ⁸	1	62	18 $\pm$ 21
Aug. 1992 ³	9	88	30 $\pm$ 31
Jun. 1993 ³	5	154	56 $\pm$ 58
Jul. 1995	0	171	27 $\pm$ 48
Aug. 1996	<1	51	30 $\pm$ 12
Average			40 $\pm$ 28

Estimates are based on the range of mat abundance observed in these cruises. Cruise minimum, maximum and average are shown. 1995–96 mat abundance data were collected in this study (see ref. 10 for cruise tracks).

Methods

Water column imaging. We used a video plankton recorder (VPR) package²² consisting of three units: a video camera and red stroboscopic light source, a video recorder and a conductivity, temperature and depth recorder (CTD) deployed as a single unit. Profiles were collected at station 3 (30°34.16' N, 157°9.30' W), station 4 (30°31.46' N, 154°19.47' W) and station 5 (30°36.72' N, 152°18.27' W) in August 1996 at 22:00, 22:00 and 13:00 local time, respectively. Video images (visualized volume, 0.08 l) were collected at 30 frames per second synchronized with the stroboscopic light. Instantaneous pressure (converted to depth), salinity and temperature were recorded by the VPR-mounted CTD and synchronized to the VPR videos. Videotape (S-VHS format) was reviewed post-cruise. Mats were identified by their distinctive morphology of intertwining diatom chains forming macroscopic aggregations up to several centimetres in size^{7,23}.

Mat characterization. Mat abundance was determined using a computer-based, quantitative image-analysis system for the depth intervals (m) of 0–20, 20–50, 50–100 and 100–150, calculated from the VPR field of view, the length of time spent within each interval and the ship speed. Mat size was determined by image analysis^{22,24}. We assumed no overlap between diver- and VPR-enumerated mats based on the volumes sampled and the mat sizes reported.

Transport calculations. For Table 1 we used only measured abundance (determined by divers and VPR) and model-generated values for N normalized transport (ϕ , 0.14 d⁻¹; ref. 4) according to the equation $\phi = F/C$, where F is net transport ($\mu\text{mol m}^{-2} \text{d}^{-1}$) and C is integrated N ($\mu\text{mol m}^{-2}$). Cruise summary results (Table 2) were computed by adding transport rates from diver-reported *Rhizosolenia* mat abundance extrapolated to 150 m (an estimated 27% of total mats found in upper 20 m; ref. 4) to the VPR-measured mat transport. The VPR mat transport for these cruises was estimated by multiplying the diver-reported transport by the diver (0–20 m):VPA (0–150 m) mat transport ratios in Table 1 (average, 4).

Received 8 July; accepted 27 November 1998.

- Villareal, T. A. & Lipschultz, F. Internal nitrate concentrations in single cells of large phytoplankton from the Sargasso Sea. *J. Phycol.* **31**, 689–696 (1995).
- Villareal, T. A., Altabet, M. A. & Culver-Rymasz, K. Nitrogen transport by vertically migrating diatom mats in the North Pacific Ocean. *Nature* **363**, 709–712 (1993).

- Villareal, T. A., Woods, S., Moore, J. K. & Culver-Rymasz, K. Vertical migration of *Rhizosolenia* mats and their significance to NO₃ fluxes in the central North Pacific gyre. *J. Plankt. Res.* **18**, 1103–1121 (1996).
- Richardson, T. L., Cullen, J. J., Kelley, D. E. & Lewis, M. R. Potential contributions of vertically migrating *Rhizosolenia* to nutrient cycling and new production in the open ocean. *J. Plankt. Res.* **20**, 219–241 (1998).
- Karl, D. et al. The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* **388**, 533–538 (1997).
- Brzezinski, M. A., Villareal, T. A. & Lipschultz, F. Silica production and the contribution of diatoms to new and primary production in the central North Pacific. *Mar. Ecol. Prog. Ser.* **167**, 89–101 (1998).
- Carpenter, E. J. et al. *Rhizosolenia* mats. *Limnol. Oceanogr.* **22**, 739–741 (1977).
- Villareal, T. A. & Carpenter, E. J. Nitrogen-fixation, suspension characteristics and chemical composition of *Rhizosolenia* mats in the central North Pacific Gyre. *Biol. Oceanogr.* **6**, 327–345 (1989).
- Banase, K. Clouds, deep chlorophyll maxima and the nutrient supply to the mixed layer of stratified water bodies. *J. Plankt. Res.* **9**, 1031–1036 (1987).
- Emerson, S. & Hayward, T. L. Chemical tracers of biological processes in shallow waters of North Pacific: preformed nitrate distribution. *J. Mar. Res.* **53**, 499–513 (1995).
- Parsons, T. R., Maita, Y. & Lalli, G. M. *A Manual of Chemical and Biological Methods for Seawater Analysis* (Pergamon, New York, 1984).
- Shipe, R. F., Brzezinski, M. A., Pilskaln, C. & Villareal, T. A. Silica production of *Rhizosolenia* mats in the central North Pacific gyre. (submitted).
- Karl, D. M. HOT and the North Pacific Ocean. *Nature* **378**, 21–22 (1995).
- Karl, D. M. & Lukas, R. The Hawaii Ocean Time-series (HOT) program: Background, rationale and field implementation. *Deep-Sea Res.* **43**, Part II, 129–156 (1996).
- Karl, D. M., Hebel, D. V., Björkman, K. & Letelier, R. M. The role of dissolved organic matter release in the productivity of the oligotrophic North Pacific Ocean. *Limnol. Oceanogr.* **43**, 1270–1286 (1998).
- Eppley, R. W. & Peterson, B. J. Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* **282**, 677–680 (1979).
- Goldman, J. C. in *Toward a Theory on Biological-Physical Interactions in the World Ocean* (ed. Rothschild, B. J.) 273–296 (Kluwer Academic, Dordrecht, 1988).
- Hayward, T. L. The shallow oxygen maximum layer and primary production. *Deep-Sea Res.* **41**, 559–574 (1994).
- Yoder, J. A., Ackleson, S. & Balch, W. M. A line in the sea. *Nature* **371**, 689–671 (1994).
- McGillcuddy, D. J. Jr & Robinson, A. R. Eddy-induced nutrient supply and new production in the Sargasso Sea. *Deep-Sea Res.* **44**, 1427–1450 (1997).
- Hall, M. M., Niiler, P. P. & Schmitz, W. J. Jr Circulation in the eastern North Pacific: results from a current meter array along 152°W. *Deep-Sea Res.* **44**, 1127–1146 (1997).
- Davis, C. S., Gallagher, S. M., Berman, M. S., Haury, L. R. & Strickler, J. R. The Video Plankton Recorder (VPR): Design and initial results. *Ergeb. Limnol.* **32**, 67–81 (1992).
- Allredge, A. L. & Silver, M. W. Abundance and production rates of floating diatom mats (*Rhizosolenia castracanei* and *Rhizosolenia imbricata* var. *shrubsolei*) in the eastern Pacific Ocean. *Mar. Biol.* **66**, 83–88 (1982).
- Pilskaln, C. H., Lehmann, C., Paduan, J. B. & Silver, M. W. Spatial and temporal dynamics in marine aggregate abundance, sinking rate, and flux: Monterey Bay, central California. *Deep-Sea Res.* (in the press).
- Martinez, L., Silver, M. W., King, J. M. & Allredge, A. L. Nitrogen fixation by floating diatom mats: a source of new nitrogen to oligotrophic ocean waters. *Science* **221**, 152–154 (1983).

Acknowledgements. We thank L. Joseph, A. Mansfield, C. L. De La Rocha, R. Shipe, W. Golden, S. Gallagher, C. Lehmann and C. Darkangelo for providing diving and technical support. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to T.A.V. (e-mail: tracy@utmsi.utexas.edu).

Strain-induced magma fragmentation in explosive eruptions

Paolo Papale

CNR-CSGSDA, via S. Maria 53, I-56126 Pisa, Italy

Explosive eruptions are the most powerful and destructive type of volcanic activity. These eruptions are characterized by magma fragmentation, the process through which a bubbly or foamy magma is transformed into a gas-pyroclast dispersion¹. Although magma fragmentation has been investigated both experimentally^{2–7} and theoretically^{8–10}, and the basic transport phenomena that occur in a volcanic conduit have been modelled^{11–17}, the underlying mechanism responsible for magma fragmentation is still poorly understood. This lack of understanding seriously limits our ability to forecast volcanic hazards, preventing reliable discrimination between conditions that lead to explosive and effusive eruptions. Here I develop a model in which a fragmentation criterion, based on a rate-limited crossing of the glass transition^{1,18–20}, is incorporated into a multiphase fluid-dynamic description of magma ascent¹⁷. The numerical results of this model demonstrate the feasibility of strain-induced brittle fragmentation of magma in volcanic eruptions, and reconcile experimental

with theoretical studies as well as with the observed volcanic products of large explosive eruptions.

Conduit flow simulations^{14,17} show a remarkably high magma acceleration in a conduit region where the volume fraction of the exsolved gas is comparable to that of natural pumice²¹. Such an acceleration is associated with large elongational strain of magma that may result in a rheological transition from ductile to brittle magma behaviour¹. A similar transition is observed when a thin fibre of viscous magma is subjected to an increasing strain rate^{18–20}. After the onset of pseudoplastic behaviour, the fibre breaks when the rate of accumulation of stress is too large compared to the rate of its dissipation by viscous flow. The fibre breakage corresponds to a rate-limited crossing of the glass transition¹.

The Maxwell relation^{18,22} provides the theoretical framework for expressing a criterion for magma breakage:

$$\frac{dv_z}{dz} > k \frac{1}{\tau} = k \frac{G_\infty}{\eta_s} \quad (1)$$

where v is velocity, z is the elongation direction ($dv_z/dz = \dot{\gamma}$ is the elongational strain rate), k is a constant, τ is the magma structural relaxation time, η_s is the newtonian mixture viscosity at zero frequency, and G_∞ is the elastic modulus at infinite frequency. The constant k in equation (1) is determined experimentally^{18,19}. For a wide range of compositions including nephelinite, tholeiite, andesite and rhyolite, brittle magma failure occurs at a strain rate two orders of magnitude below the critical strain rate equal to the inverse of the structural relaxation time, corresponding to $k = 0.01$. G_∞ is found to be scarcely dependent on magma composition and temperature, and is in the range 3–30 GPa (ref. 18).

An extension of equation (1) to magma fragmentation in volcanic conduits implies the assumption that the multiphase magma below, and at, fragmentation behaves as a pseudofluid with homogeneous properties equal to the mixture properties. This is supported by calculated negligible mechanical interphase disequilibrium along the entire conduit length (except very close to the conduit exit¹⁷). In addition, it is also assumed that the multiphase nature of magma does not result in large differences of the values of k and G_∞ with respect to the values determined for pure liquid magma. This assumption requires future experimental work in order to be evaluated.

The fragmentation criterion expressed by equation (1) is included in the one-dimensional, multiphase non-equilibrium magma ascent modelling¹⁷. The fluid-dynamic model is an evolution of previous models^{12–14}, and accounts for choking of magma at the conduit exit, friction at the conduit walls, interphase drag, magma composition-dependent density, viscosity and volatile saturation surface, a two-component $H_2O + CO_2$ exsolving gas phase, gas-exsolution-induced variations of magma properties, and the effect of crystals and gas bubbles on the mixture viscosity. Several magma ascent simulations have been carried out in order to cover a representative spectrum of eruption conditions in terms of liquid magma composition, crystal and volatile content, conduit size, magma temperature, mass flow-rate, and initial pressure in the magma chamber (Table 1). An average value of G_∞ equal to 25 GPa has been used¹⁹, after verifying that a one order of magnitude difference in the product kG_∞ produces small changes (<10% for most flow variables) in the calculated conditions at fragmentation.

In Table 1, the run names beginning with R, D and T refer to rhyolitic, dacitic and tholeiitic magma at $T = 1,100, 1,150$ and $1,470$ K, respectively. D and L are the conduit diameter and length, while $\dot{m}$, α , P , u , z and $\dot{\gamma}$ are the calculated mass flow-rate, gas volume fraction, pressure, velocity, vertical coordinate (with origin at the conduit base and upward positive) and elongational strain rate, respectively. The superscript F refers to the conditions at fragmentation. The stagnation pressure in the magma chamber is equal to 192, 110 and 55 MPa for conduit lengths of 7, 4 and 2 km, respectively. The column labelled 'e.b.c.' indicates the assumed boundary condition at the conduit exit, where 'c.f.' and 's.f.' stand for critical and subcritical flow, respectively; the column labelled 'Fragm.' indicates the adopted fragmentation criterion.

Figure 1 shows the calculated distribution of gas volume fraction and non-dimensional pressure (Fig. 1a), non-dimensional velocity (Fig. 1b), mixture viscosity (Fig. 1c), elongational strain rate and inverse of the structural relaxation time (Fig. 1d) for three simulations pertaining to rhyolitic (R2), dacitic (D3) and tholeiitic (T4) magma composition for which equation (1) has been used as a fragmentation criterion. In addition, the distributions pertaining to the same conditions of run T4 but assuming that fragmentation occurs when the critical gas volume fraction of 0.75 is reached^{11,23} are

Table 1 Input data and results of simulations

Run	Input data							Results					
	D (m)	L (m)	H_2O (wt%)	CO_2 (wt%)	Cryst. (vol.%)	Fragm.	e.b.c.	$\dot{m}$ (kg s ⁻¹)	α^F	P^F (MPa)	u^F (m s ⁻¹)	z^F (m)	$\dot{\gamma}^F$ (s ⁻¹)
R1	100	7,000	6.0	0	0	Eq. (1)	c.f.	2.58×10^8	0.757	18.9	55.5	4,913.9	9.66
R2	100	7,000	4.0	0	0	Eq. (1)	c.f.	2.14×10^8	0.701	17.4	35.2	5,113.4	6.76
R3	100	7,000	3.0	0	0	Eq. (1)	c.f.	1.45×10^8	0.667	13.1	23.3	5,038.5	4.89
R4	100	7,000	2.0	0	0	Eq. (1)	c.f.	0.56×10^8	0.642	9.33	8.37	4,873.4	2.08
R5	100	7,000	4.0	1.0	0	Eq. (1)	c.f.	2.02×10^8	0.701	15.7	35.7	5,020.8	15.2
R6	100	7,000	4.0	0	0	$\alpha = 0.75$	c.f.	2.63×10^8	0.750*	19.5	54.9	4,893.0	6.13
R7	100	4,000	4.0	0	0	Eq. (1)	c.f.	1.91×10^8	0.703	15.3	34.1	4,972.5	6.36
R8	100	7,000	3.0	0	0	$\alpha = 0.75$	c.f.	1.22×10^8	0.750*	9.48	25.9	5,301.6	0.21
D1	100	7,000	6.0	0	0	Eq. (1)	c.f.	3.61×10^8	0.830	15.9	95.5	6,432.8	23.4
D2	100	7,000	4.0	0	0	Eq. (1)	c.f.	2.89×10^8	0.805	10.8	75.7	6,351.5	18.3
D3	100	7,000	2.0	0	0	Eq. (1)	c.f.	1.08×10^8	0.804	5.22	28.4	6,078.1	9.20
D4	100	7,000	1.5	0	0	Eq. (1)	c.f.	0.58×10^8	0.823	3.44	16.9	6,024.7	8.37
D5	100	7,000	4.0	0.5	0	Eq. (1)	c.f.	2.58×10^8	0.813	10.8	70.3	6,176.7	15.7
D6	100	7,000	4.0	1.0	0	Eq. (1)	c.f.	2.36×10^8	0.822	11.0	66.6	5,992.3	15.2
D7	60	7,000	4.0	0	0	Eq. (1)	c.f.	0.71×10^8	0.802	11.0	50.8	5,763.4	18.3
D8	100	4,000	4.0	0	0	Eq. (1)	c.f.	2.06×10^8	0.824	9.52	59.7	5,263.6	14.9
D9	100	7,000	4.0	0	18	Eq. (1)	c.f.	2.07×10^8	0.764	11.3	43.2	5,736.3	9.88
D10	100	7,000	4.0	0	35	Eq. (1)	c.f.	1.23×10^8	0.696	12.4	19.1	4,967.0	3.89
D11	100	7,000	4.0	0	45	Eq. (1)	c.f.	0.70×10^8	0.628	13.8	8.76	4,360.7	1.53
T1	10	2,000	2.0	0	16	Eq. (1)	s.f.	5.00×10^4	-	-	-	-	-
T2	8	2,000	1.0	0	16	Eq. (1)	s.f.	5.25×10^3	-	-	-	-	-
T3	4	2,000	2.0	0	16	Eq. (1)	s.f.	1.07×10^3	-	-	-	-	-
T4	4	2,000	2.0	0	16	Eq. (1)	c.f.	3.50×10^5	-	-	-	-	-
T5	4	2,000	2.0	0	16	$\alpha = 0.75$	c.f.	4.06×10^5	0.750*	7.74	43.2	6,485.7	0.19
T6	10	7,000	2.0	0	50	Eq. (1)	c.f.	2.94×10^8	0.938	1.46	21.5	6,356.1	73.3

*Fragm.: indicates the adopted fragmentation criterion; superscript F refers to the conditions of fragmentation; 'e.b.c.' shows the assumed boundary condition at the conduit exit (c.f. and s.f. represent critical flow and subcritical flow, respectively). Variables are defined in the text.

* Assumed in the simulation.

also shown (run T5). In the first two cases (R2 and D3) strain-induced magma fragmentation is predicted to occur, following the development of large gradients of gas volume fraction, pressure, velocity and mixture viscosity. The main processes leading to the establishment of such large gradients, and finally to magma fragmentation, are discussed below.

Along the conduit the magmatic pressure decreases, and eventually the gas nucleation level is reached. Gas exsolution results in a mixture viscosity increase that translates into larger friction at the conduit walls and further pressure decrease, leading to a sort of feedback through which both pressure and gas volume fraction change more and more rapidly. As a result, the magma is forced to accelerate to satisfy mass conservation. The overall result is an increase in both elongational strain rate and structural relaxation

time (the latter is related (equation (1)) to rapid mixture viscosity increase) that finally pushes the magma into a region of brittle response beyond the glass transition and culminates in magma fragmentation. In the case of run T4, the low viscosity of tholeiitic magma (implying small structural relaxation time) does not allow crossing of the glass transition. In this case, and in runs T1–T3, the magmatic mixture leaves the vent as a liquid continuum with dispersed gas bubbles and gives rise to an effusive eruption, with a calculated exit gas volume fraction in the range 0.1–0.9 depending on the eruption conditions. If, however, a critical gas volume fraction of 0.75 is used as a fragmentation criterion, the same conditions of run T4 produce magma fragmentation and a much more efficient acceleration close to the conduit exit (run T5). Strain-induced fragmentation of tholeiitic magma is only possible for unusual conditions producing very large viscosity and/or strain rate—as in run T6 where a crystal content of 50 vol.% is associated with an increase in viscosity of three orders of magnitude²⁴. In this case the predicted gas volume fraction at fragmentation is comparable to that of natural reticulite²⁵ (Table 1).

The calculated range of gas volume fraction at fragmentation (Table 1) closely corresponds to the range of measured vesicularity in juvenile magmatic clasts from many eruptions all around the world^{21,25–28}. It is remarkable that the most common conditions for explosive eruptions among those reported in Table 1—composition from dacite to rhyolite, total volatile content between 4 and 6 wt%, and crystal content from 0 to 30 vol.%—are associated with a gas volume fraction at fragmentation in the narrow range 0.7–0.83, corresponding to the peak of frequency of measured pumice vesicularities^{26,27}. Furthermore, the fragmentation criterion in equation (1) provides a simple explanation for the observed inverse relationship between pumice vesicularity and inferred magma viscosity at fragmentation^{21,29} (Fig. 2). Indeed, a larger magma viscosity implied a larger structural relaxation time and a more rapid pressure decrease due to dissipation by wall friction, resulting in magma fragmentation at a lower gas volume fraction and thus in less vesicular magma fragments.

The present work demonstrates that strain-induced magma fragmentation in volcanic conduits is a feasible process. Other proposed fragmentation mechanisms including gas bubble overpressure exceeding the magma tensile strength, propagation of a decompression wave, and thin film instability and permeability development in a magmatic foam (see the review by Dingwell¹) can not be ruled out. In particular, the first of these other mechanisms has been evaluated by numerical simulation of volcanic conduit flow¹⁰. It should be noted that both gas bubble overpressure and strain criteria for magma fragmentation are essentially of the form $\eta\dot{\gamma} > X$, where X is a quantity with the dimension of pressure that

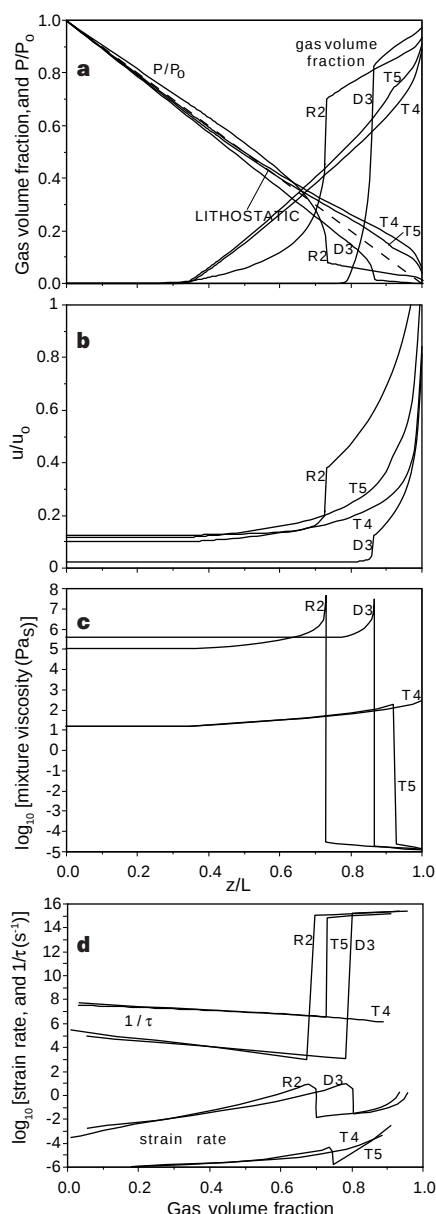


Figure 1 Calculated flow variables along the volcanic conduit. **a**, Gas volume fraction and non-dimensional pressure (P is pressure, and the curve labelled "lithostatic" denotes the lithostatic pressure distribution); **b**, non-dimensional velocity; **c**, mixture viscosity; **d**, elongational strain rate and inverse of the structural relaxation time τ , for simulations R2, D3, T4 and T5 in Table 1. The reference pressure P_0 in **a** corresponds to the stagnation pressure (see text). The reference velocity u_0 in **b** is 100 m s^{-1} , except for run D3 when it is 140 m s^{-1} .

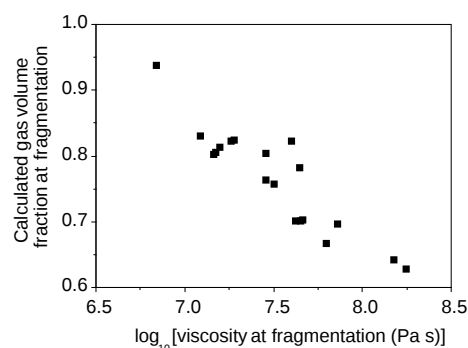


Figure 2 Calculated fragmentation conditions. Shown is the gas volume fraction at fragmentation plotted against the calculated magma viscosity at fragmentation for the simulations in Table 1 for which the fragmentation criterion in equation (1) has been used. An inverse relation is found, with a small dispersion of points reflecting a secondary dependence of the fragmentation conditions on factors other than magma viscosity.

embodies the details of the fragmentation mechanism and the assumptions and simplifications concerning magma rheology and degassing. As η and $\dot{\gamma}$ are predicted to increase very rapidly in a narrow conduit region in response to magma degassing (Fig. 1), relatively large variations in the value of X do not lead to significant displacements of the calculated fragmentation level, and the two mechanisms show good agreement in the predicted fragmentation conditions.

The model reported here makes predictions that are in line with experiments and observations, as follows: the consistency between the experimental conditions for brittle magma fragmentation and the fluid dynamics of magma ascent; the calculated range of gas volume fraction at fragmentation in close agreement with measured pumice vesicularities; the inverse relationship between pumice vesicularity and magma viscosity (which is observed in natural samples and suggested by the modelling results); and the discrimination between high-viscosity (mainly explosive) and low-viscosity (mainly effusive) eruptions. This suggests that the proposed criterion for magma fragmentation could be operating during sustained explosive eruptions. □

Received 2 March; accepted 9 October 1998.

- Dingwell, D. B. Volcanic dilemma: flow or blow? *Science* **273**, 1054–1055 (1996).
- Mader, H. M. *et al.* Experimental simulations of explosive degassing of magma. *Nature* **372**, 85–88 (1994).
- Phillips, J. C., Lane, S. J., Lejeune, A. M. & Hilton, M. The gum rosin-acetone system as an analogue to the degassing behaviour of hydrated magmas. *Bull. Volcanol.* **57**, 263–268 (1995).
- Mader, H. M., Phillips, J. C., Sparks, R. S. J. & Sturtevant, B. Dynamics of explosive degassing of magma: observations of fragmenting two-phase flows. *J. Geophys. Res.* **101**, 5547–5560 (1996).
- Alidivirov, M. & Dingwell, D. B. Magma fragmentation by rapid decompression. *Nature* **380**, 146–148 (1996).
- Zimanowski, R. B., Büttner, R., Lorentz, V. & Häfele, H. G. Fragmentation of basaltic melt in the course of explosive volcanism. *J. Geophys. Res.* **102**, 803–814 (1997).
- Zhang, Y., Sturtevant, B. & Stolper, E. M. Dynamics of gas-driven eruptions: experimental simulations using CO₂-H₂O-polymer system. *J. Geophys. Res.* **102**, 3077–3096 (1997).
- Wohletz, K. H., Sheridan, M. F. & Brown, W. K. Particle size distributions and sequential fragmentation/transport theory applied to volcanic ash. *J. Geophys. Res.* **94**, 15703–15721 (1989).
- Alidivirov, M. A. A model for viscous magma fragmentation during volcanic blasts. *Bull. Volcanol.* **56**, 459–465 (1994).
- Barmin, A. A. & Mel'nik, O. É. Eruption dynamics of high-viscosity gas-saturated magmas. *Izv. Ros. Akad. Nauk. Mekh. Zhidk. Gaza* **2**, 49–60 (1993).
- Wilson, L., Sparks, R. S. J. & Walker, G. P. L. Explosive volcanic eruptions—IV. The control of magma properties and conduit geometry on eruption column behaviour. *Geophys. J. R. Astron. Soc.* **63**, 117–148 (1980).
- Dobran, F. Nonequilibrium flow in volcanic conduits and application to the eruptions of Mt. St. Helens on May 18, 1980, and Vesuvius in AD 79. *J. Volcanol. Geotherm. Res.* **49**, 285–311 (1992).
- Papale, P., Dobran, F. Modelling of the ascent of magma during the plinian eruption of Vesuvius in AD 79. *J. Volcanol. Geotherm. Res.* **58**, 101–132 (1993).
- Papale, P., Dobran, F. Magma flow along the volcanic conduit during the Plinian and pyroclastic flow phases of the May 18, 1980, Mount St. Helens eruption. *J. Geophys. Res.* **99**, 4355–4373 (1994).
- Ramos, J. I. One-dimensional, time-dependent, homogeneous, two-phase flow in volcanic conduits. *Int. J. Num. Methods Fluids* **21**, 253–278 (1995).
- Woods, A. W. & Bower, S. M. The decompression of volcanic jets in a crater during explosive volcanic eruptions. *Earth Planet. Sci. Lett.* **131**, 189–205 (1995).
- Papale, P. From magma to tephra: Modeling physical processes of explosive volcanic eruptions (eds Freundt, A. & Rosi, M.) *Developments in Volcanology* (Elsevier, Amsterdam, in the press).
- Dingwell, D. B. & Webb, S. L. Structural relaxation in silicate melts and non-Newtonian melt rheology in geologic processes. *Phys. Chem. Miner.* **16**, 508–516 (1989).
- Webb, S. L. & Dingwell, D. B. Non-Newtonian rheology of igneous melts at high stresses and strain rates: experimental results for rhyolite, andesite, basalt, and nephelinite. *J. Geophys. Res.* **95**, 15695–15701 (1990).
- Dingwell, D. B. The brittle-ductile transition in high-level granitic magmas: material constraints. *J. Petrol.* **38**, 1635–1644 (1997).
- Thomas, N., Jaupart, C. & Vergnolle, S. On the vesicularity of pumice. *J. Geophys. Res.* **99**, 15633–15644 (1994).
- Maxwell, J. C. On the dynamical theory of gases. *Phil. Trans. R. Soc. Lond.* **157**, 49–88 (1867).
- Sparks, R. S. J. The dynamics of bubble formation and growth in magmas: a review and analysis. *J. Volcanol. Geotherm. Res.* **3**, 1–37 (1978).
- Lejeune, A.-M. & Richet, P. Rheology of crystal-bearing silicate melts: an experimental study at high viscosities. *J. Geophys. Res.* **100**, 4215–4229 (1995).
- Mangan, M. T. & Cashman, K. V. The structure of basaltic scoria and reticulate and inferences for vesiculation, foam formation, and fragmentation in lava fountains. *J. Volcanol. Geotherm. Res.* **73**, 1–18 (1996).
- Whitham, A. G. & Sparks, R. S. J. Pumice. *Bull. Volcanol.* **48**, 209–223 (1986).
- Houghton, B. F. & Wilson, C. J. N. A vesicularity index for pyroclastic deposits. *Bull. Volcanol.* **51**, 451–462 (1989).
- Klug, C. & Cashman, K. V. Permeability development in vesiculating magmas: implications for fragmentation. *Bull. Volcanol.* **58**, 87–100 (1996).
- Kaminsky, E. & Jaupart, C. Expansion and quenching of vesicular magma fragments in Plinian eruptions. *J. Geophys. Res.* **102**, 12187–12203 (1997).

Acknowledgements. This work was supported by the Italian Gruppo Nazionale per la Vulcanologia, and performed when the author was a member of the Volcanic Simulation Group of Pisa.

Correspondence and requests for materials should be addressed to the author (e-mail: papale@dst.unipi.it).

A neuronal representation of the location of nearby sounds

Michael S. A. Graziano, Lina A. J. Reiss & Charles G. Gross

Department of Psychology, Princeton University, Princeton, New Jersey 08544, USA

Humans can accurately perceive the location of a sound source—not only the direction, but also the distance^{1–9}. Sounds near the head, within ducking or reaching distance, have a special saliency. However, little is known about this perception of auditory distance. The direction to a sound source can be determined by interaural differences, and the mechanisms of direction perception have been studied intensively¹; but except for studies on echolocation in the bat¹⁰, little is known about how neurons encode information on auditory distance. Here we describe neurons in the brain of macaque monkeys (*Macaca fascicularis*) that represent the auditory space surrounding the head, within roughly 30 cm. These neurons, which are located in the ventral premotor cortex, have spatial receptive fields that extend a limited distance outward from the head.

The ventral premotor cortex (PMv) is a multimodal, sensory-motor area located in the frontal lobe just anterior to primary motor cortex. Most PMv neurons respond to touch, and about 40% also respond to visual stimuli^{11–16}. For these bimodal neurons, the visual receptive field extends from the region of the tactile receptive field into the space immediately adjacent (Fig. 1a). These neurons usually do not respond to distant visual stimuli, more than 30 cm from the tactile receptive field. That is, they represent the space near the body, within the monkey's reach. Here we studied neurons in PMv whose tactile receptive fields included the back of the monkey's head, and found that 53% were trimodal, responding to tactile, visual and auditory stimuli (Table 1). The mean auditory response latency was 46 ms (s.d. = 10.9 ms).

Figure 1b shows the responses of a typical trimodal neuron. The tactile receptive field (stippled) covered the contralateral side of the head. The neuron responded to visual stimuli within about 20 cm of the contralateral side of the face. In addition to these tactile and visual responses, the neuron also responded to sounds produced near the contralateral side of the head. Jangling keys, claps, crinkling paper, and bursts of white noise all gave strong responses. Sine waves of various frequencies (200–3,200 Hz) did not elicit responses. To test the auditory receptive field, six speakers were arranged around the head at the azimuth angles shown (Fig. 1b) and at a distance of 10 cm, and bursts of white noise were presented for 0.3 s. The neuron responded to all six speaker positions, but preferred position 2, near the contralateral side of the face (Fig. 1b). That is, the auditory receptive field matched the location of the visual and tactile receptive fields, although the tactile receptive field was larger and also included the back of the head. The calculated preferred direction of the auditory response for this neuron was 27° contralateral to the direction that the monkey was facing.

Figure 1c shows the results for 43 trimodal neurons. Each arrow

Table 1 Number of neurons in the different response categories

	Tactile response on back of head	No tactile response on back of head
Unresponsive	—	70
Visual	—	12
Auditory	—	1
Tactile	10	51
Tactile + visual	67	64
Tactile + visual + auditory	88	3
Total	165	201

represents the preferred auditory direction for one neuron. The auditory responses in PMv clearly span the entire contralateral space, and seem to represent the space in front of the head (31 neurons) more densely than the space behind the head (12 neurons; $\chi^2 = 8.4$, $P < 0.01$). Almost all neurons with an auditory response also responded to a touch on the ears or back of the head (Table 1). Yet the tactile receptive fields of these trimodal neurons were rarely restricted to the back of the head (9/91, 10%); most also included parts of the cheek, eyebrow, snout or jaw (82/91, 90%). Likewise, the visual receptive fields were sometimes located in the periphery and sometimes extended into the centre of the visual field. That is, the trimodal neurons in PMv do not form a back-of-the-head representation, but instead form a complete representation of the space around the head. As already described, the visual representation in PMv emphasizes the space within reaching distance. Do the auditory receptive fields have a similar spatial limit?

Figure 2 shows data from four trimodal neurons. We manipulated both the amplitude of the sound and the distance between the

speaker and the head. In Fig. 2a, the neuron responded to sounds presented 10 cm from the head. Sounds presented 30 or 50 cm away did not elicit a response, even though they covered the same range of amplitudes measured at the head. The effect of distance on the response of the neuron was statistically significant, but there was no significant overall effect of the amplitude of the sound on the response of the neuron (see regression analysis in legend of Fig. 2). Data from another neuron are shown in Fig. 2b. Like the cell in Fig. 2a, this neuron responded significantly better to closer stimuli, but unlike that cell, it responded significantly better to higher-amplitude stimuli as well. The neuron in Fig. 2c had an inhibitory response to auditory stimuli and an excitatory response to tactile stimuli. It responded best, that is, had the lowest firing rate, to sounds presented 10 cm away, and responded less well to sounds at 30 or 50 cm. Again the effect both of distance and of amplitude was significant. We often found trimodal neurons with an inhibitory auditory response and an excitatory tactile response, or vice versa (17 of 91 trimodal neurons, 19%). This result demonstrates that the response to nearby sounds is not caused by the sound mechanically stimulating the tactile receptive field. Finally, for the neuron in Fig. 2d, the response showed no significant dependence on the distance to the sound source, and instead depended on the amplitude of the sound.

In total, 44 neurons (34 in monkey 1, 10 in monkey 2) were tested for dependence on stimulus amplitude and distance. Of these, 15

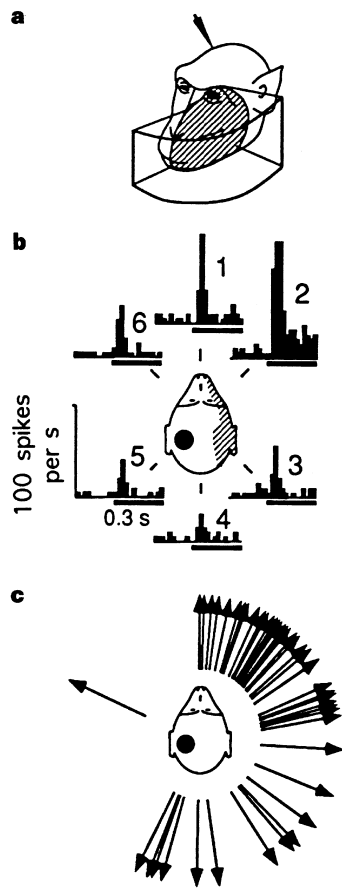


Figure 1 Responses of bimodal and trimodal neurons in PMv. **a**, Receptive fields of a typical bimodal, visual-tactile neuron. The tactile receptive field (shaded) is on the front of the face contralateral to the recording electrode (indicated by the arrowhead). The visual receptive field (boxed) is confined to a region of space within about 10 cm of the tactile receptive field. **b**, Responses of a typical trimodal, visual-tactile-auditory neuron. The tactile receptive field is contralateral to the recording electrode (indicated by the black spot) and includes the ear and back of the head. The visual receptive field (not shown) extends about 20 cm into the space near the contralateral side of the face. The histograms show the response, summed over ten trials, to a burst of white noise presented 10 cm away at the indicated azimuth angles. **c**, The calculated preferred direction of the auditory response for 43 trimodal neurons. Each arrow shows the result for one neuron. Preferred direction is given by $(\Sigma \phi_n r_n) / \Sigma r_n$, where ϕ_n = the angular position of speaker n , and r_n = the neuron's response to speaker n (mean number of spikes per s in the stimulus period).

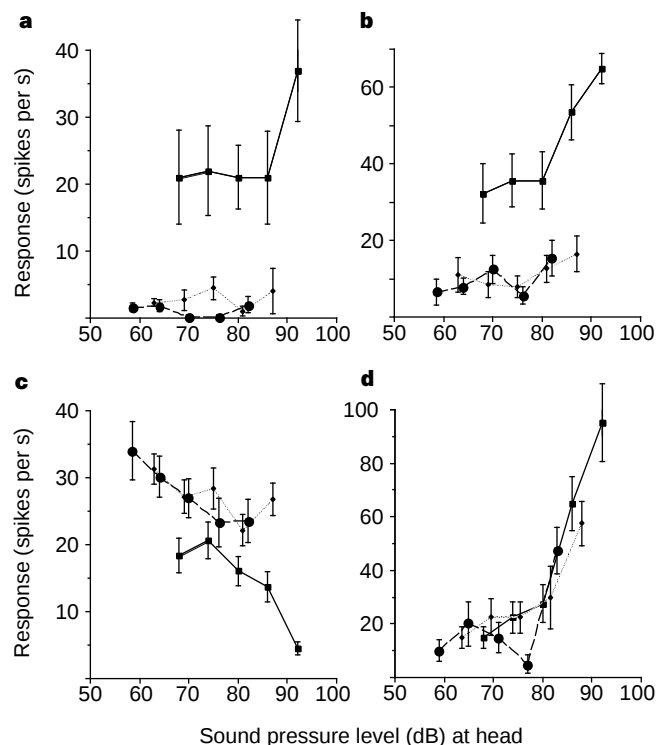


Figure 2 Auditory responses of four trimodal neurons to white noise, presented at five amplitudes and three distances from the head. x-axis, sound pressure level of stimulus measured at the head; y-axis, response of neuron in spikes per s. Distance of speaker from the head: squares, 10 cm; diamonds, 30 cm; circles, 50 cm. Each point is based on ten trials. Error bars, standard error of the mean. The neurons in **a** and **b** responded best to the closest stimuli. The neuron in **c** had an inhibitory response; it responded better (lower firing rate) to closer stimuli. In **d**, the neuron responded better to higher-amplitude stimuli, but showed no sensitivity to the distance of the stimulus. A linear regression analysis²⁶ was done to test the significance of the effect of distance (t_d) and of amplitude (t_a) independently of each other. For the neuron in **a**, $t_d = -7.194$, $P < 0.001$; $t_a = 1.335$, $P = 0.184$; **b**, $t_d = -7.235$, $P < 0.001$; $t_a = 3.711$, $P < 0.001$; **c**, $t_d = 4.225$, $P < 0.001$; $t_a = -4.745$, $P < 0.001$; **d**, $t_d = -1.241$, $P = 0.217$; $t_a = 7.549$, $P < 0.001$.

(34%) responded significantly better to closer stimuli but were unaffected by the amplitude of the stimulus; 15 (34%) responded significantly better to higher-amplitude stimuli but were unaffected by the distance to the stimulus; and 11 (25%) responded significantly better to closer stimuli and, independently, to higher-amplitude stimuli. Amplitude is one of many possible cues that humans use to determine the distance to an auditory event¹⁷. Therefore, we suggest that the amplitude-sensitive neurons described here use this particular cue to code distance. These neurons will tend to respond to nearby stimuli because they respond better to higher-amplitude sounds. However, more than half of the neurons (59%) code distance by means of some other cue or combination of cues, such that they respond to nearby stimuli independently of the amplitude. Reverberation of the sound from the walls of the room may be important¹⁸. Another possible set of cues for distance involves familiarity with the sound source¹⁹. However, the first neuron tested in monkey 2 was significantly dependent on distance even though the monkey had never heard the stimulus before. Another possible cue is the difference in amplitude between the two ears; a very large difference implies a sound source close to one ear. However, the neurons were sensitive to distance even when the stimulus was presented on the midline, that is, when the amplitude was equal in both ears. Finally, the calculation of distance near the head may depend on the highly complex distorting effect of the head and pinnae on the sound spectrum¹. This last effect would be especially pronounced at such close distances as 10 cm. A full analysis of the relative influence of these different cues will require further experiments.

The cortical pathways for auditory spatial processing are not well understood. Perhaps distance information is calculated in a different brain area and then relayed to the trimodal neurons in PMv. Recently, we studied neurons in a portion of parietal area 7b²⁰, in the upper bank of the later sulcus, and found similar trimodal, tactile–visual–auditory neurons (M.S.A.G. and C.G.G., manuscript in preparation). Area 7b projects to PMv^{21,22}, but whether the trimodal region of 7b projects to the trimodal region of PMv has not yet been determined.

Previous experiments showed that multimodal neurons in PMv encode the locations of nearby objects, within about reaching distance, through touch, vision, and even visual memory^{11–16}. Our results show that PMv neurons also represent nearby auditory space. Because a high proportion of PMv neurons respond during movements of the head, mouth, arms and hands, the purpose of this multimodal map of space may be to guide movements towards and around the objects that surround the body^{23,24}. □

Methods

Two adult *M. fascicularis* were trained to sit in a primate chair; they did not perform any task. (For details of the experimental procedures, see ref. 16.) During daily recording sessions, a microdrive was used to lower an electrode into PMv. Once a neuron was isolated, it was tested for somatosensory, visual and auditory responses. Somatosensory receptive fields were plotted by manipulating the joints and stroking the skin. Visual receptive fields were plotted with objects presented on a wand. Auditory stimuli included tones, clicks, claps, jingling keys and other sounds. Controlled tests were done using white noise (20–22,000 Hz) presented over Cambridge Soundworks 3-inch (76.2 mm) speakers mounted in a circular array around the monkey's head at ear level. The angular position and distance of the speakers to the head was adjustable. The sound pressure level of the stimuli was measured at the monkey's head using a Radio Shack sound level meter, repeatedly calibrated with a 0.25-inch (6.35 mm) Bruel and Kjaer microphone. Neurons were tested either with the speaker behind the head, or in the dark, so that the monkey could not see the distance to the sound source. Eye position was not controlled during the presentation of auditory stimuli. Some PMv neurons are influenced by eye position^{14,16,25}. However, the short latency of the auditory response eliminates the possibility that it was caused by a change in eye position elicited by the presentation of the stimulus. In addition, there are no reports of

transient bursts of activity in PMv associated with eye movement, whereas most of the auditory responses in PMv were transient, short-latency bursts (Fig. 1b).

Received 29 September; accepted 23 November 1998.

1. Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localization* (transl. Allen, J. S.) (MIT Press, Cambridge, Massachusetts, 1997).
2. Clifton, R. K., Rochat, P., Robin, D. J. & Berthier, N. E. Multimodal perception in the control of infant reaching. *J. Exp. Psychol. Hum. Percept. Perform.* **20**, 876–886 (1994).
3. Coleman, P. D. An analysis of cues to auditory depth perception in free space. *Psychol. Bull.* **60**, 302–315 (1963).
4. Coleman, P. D. Dual role of frequency spectrum in determination of auditory distance. *J. Acoust. Soc. Am.* **44**, 631–632 (1968).
5. Edwards, A. A. Accuracy of auditory depth perception. *J. Gen. Psychol.* **52**, 327–329 (1955).
6. Gamble, E. A. Intensity as a criterion in estimating the distance of sounds. *Psychol. Rev.* **16**, 416–426 (1909).
7. Gardner, M. B. Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space. *J. Acoust. Soc. Am.* **45**, 47–53 (1969).
8. Mershon, D. H. & Bowers, J. N. Absolute and relative cues for the auditory perception of egocentric distance. *Perception* **8**, 311–322 (1979).
9. von Békésy, G. *Experiments in Hearing* (McGraw-Hill, New York, 1960).
10. Suga, N. & O'Neill, W. E. Neural axis representing target range in the auditory cortex of the mustache bat. *Science* **206**, 351–353 (1979).
11. Gentilucci, M. et al. Functional organization of inferior area 6 in the macaque monkey. I. Somatotopy and the control of proximal movements. *Exp. Brain Res.* **71**, 475–490 (1988).
12. Fogassi, L. et al. Coding of peripersonal space in inferior premotor cortex (area F4). *J. Neurophysiol.* **76**, 141–157 (1996).
13. Graziano, M. S. A., Yap, G. S. & Gross, C. G. Coding of visual space by pre-motor neurons. *Science* **266**, 1054–1057 (1994).
14. Graziano, M. S. A., Hu, X. & Gross, C. G. Coding the locations of objects in the dark. *Science* **277**, 239–241 (1997).
15. Rizzolatti, G. et al. Afferent properties of periarculate neurons in macaque monkeys. II. Visual responses. *Behav. Brain Res.* **2**, 147–163 (1981).
16. Graziano, M. S. A., Hu, X. & Gross, C. G. Visuo-spatial properties of ventral premotor cortex. *J. Neurophysiol.* **77**, 2268–2292 (1997).
17. Ashmead, D. H., LeRoy, D. & Odom, R. D. Perception of the relative distances of nearby sound sources. *Percept. Psychophys.* **47**, 326–331 (1990).
18. Mershon, D. H. & King, L. E. Intensity and reverberation as factors in the auditory perception of egocentric distance. *Percept. Psychophys.* **18**, 409–415 (1975).
19. Coleman, P. D. Failure to localize the source distance of an unfamiliar sound. *J. Acoust. Soc. Am.* **34**, 345–346 (1962).
20. Graziano, M. S. A., Fernandez, T. & Gross, C. G. A comparison of bimodal, visual-tactile neurons in parietal area 7b and ventral premotor cortex of the monkey brain. *Neurosci. Abs.* **22**, 398 (1996).
21. Cavada, C. & Goldman-Rakic, P. S. Posterior parietal cortex in rhesus monkey: II: Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *J. Comp. Neurol.* **287**, 422–445 (1989).
22. Matelli, M., Camarda, R., Glickstein, M. & Rizzolatti, G. Afferent and efferent projections of the inferior area 6 in the macaque monkey. *J. Comp. Neurol.* **255**, 281–298 (1986).
23. Graziano, M. S. A. & Gross, C. G. Spatial maps for the control of movement. *Curr. Opin. Neurobiol.* **8**, 195–201 (1998).
24. Rizzolatti, G., Fadiga, L., Fogassi, L. & Gallese, V. The space around us. *Science* **277**, 190–191 (1997).
25. Boussaoud, D., Barth, T. M. & Wise, S. P. Effects of gaze on apparent visual responses of frontal cortex neurons. *Exp. Brain Res.* **93**, 423–434 (1993).
26. Cohen, J. & P. Cohen, P. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1983).

Acknowledgements. We thank E. Olson, X. Hu, S. Alisharan, M. E. Wheeler and V. Gomez for their help during the experiment.

Correspondence and requests for materials should be addressed to M.S.A.G. (e-mail: graziano@princeton.edu).

Perception's shadow: long-distance synchronization of human brain activity

Eugenio Rodriguez, Nathalie George, Jean-Philippe Lachaux, Jacques Martinerie, Bernard Renault & Francisco J. Varela

Laboratoire de Neurosciences Cognitives et Imagerie Cérébrale (LENA), CNRS UPR 640, Hôpital de la Salpêtrière, 47 Boulevard de l'Hôpital, 75651 Paris Cedex 13, France

Transient periods of synchronization of oscillating neuronal discharges in the frequency range 30–80 Hz (gamma oscillations) have been proposed to act as an integrative mechanism that may bring a widely distributed set of neurons together into a coherent ensemble that underlies a cognitive act^{1–4}. Results of several experiments in animals provide support for this idea (see, for example, refs 4–10). In humans, gamma oscillations have been

described both on the scalp^{11–16} (measured by electroencephalography and magnetoencephalography) and in intracortical recordings¹⁷, but no direct participation of synchrony in a cognitive task has been demonstrated so far. Here we record electrical brain activity from subjects who are viewing ambiguous visual stimuli (perceived either as faces or as meaningless shapes). We show for the first time, to our knowledge, that only face perception induces a long-distance pattern of synchronization, corresponding to the moment of perception itself and to the ensuing motor response. A period of strong desynchronization marks the transition between the moment of perception and the motor response. We suggest that this desynchronization reflects a process of active uncoupling of the underlying neural ensembles that is necessary to proceed from one cognitive state to another³.

Ten subjects were shown 'Mooney' faces (Fig. 1a, b), which are easily categorized as faces when presented in upright orientation, but usually seen as meaningless black and white shapes when presented upside-down^{18,19}. Subjects were asked to report as quickly as possible whether they had seen a face or not by pressing on one of two different keys. On average, $79 \pm 2\%$ of upright presentations were perceived as faces. Conversely, $76 \pm 2\%$ of upside-down presentations were reported as meaningless. We analysed only the cases of upright presentations that were perceived as faces and

inverted presentations that were perceived as meaningless, referred to here as the 'perception' and 'no-perception' conditions. The electroencephalogram (EEG) was recorded through 30 electrodes, and a precise time–frequency analysis was carried out up to 100 Hz.

We first computed the pseudo Wigner–Ville time–frequency transforms²⁰ of single trials and averaged these transforms over all trials. This procedure is best adapted to detect the so-called 'induced' gamma response, which is triggered by, but not phase-locked to, stimulus onset¹⁵ ('phase-locking' is synchronization of oscillations). We obtained two induced gamma-activity peaks (Fig. 1c, d). The first induced response, lying at 36 ± 3 Hz for both conditions, peaked at ~ 230 ms after stimulus onset, was significantly larger for the perception condition (Wilcoxon $T = 5$, $Z = 2.29$, $P < 0.05$) and has been consistently described as the correlate of the perceptual process itself^{11–17}. Similar conclusions are reached by studies of evoked potential²¹. In contrast, the second induced gamma peak has not been reported so far. It peaked at ~ 800 ms with a maximum frequency of 40 ± 5 Hz, following the subject's reaction time closely, and was slightly (but not significantly) stronger in the no-perception than in the perception condition. The latency of this peak indicates the possible involvement of post-perceptual processes.

We then studied phase synchrony, the main focus of our work. We

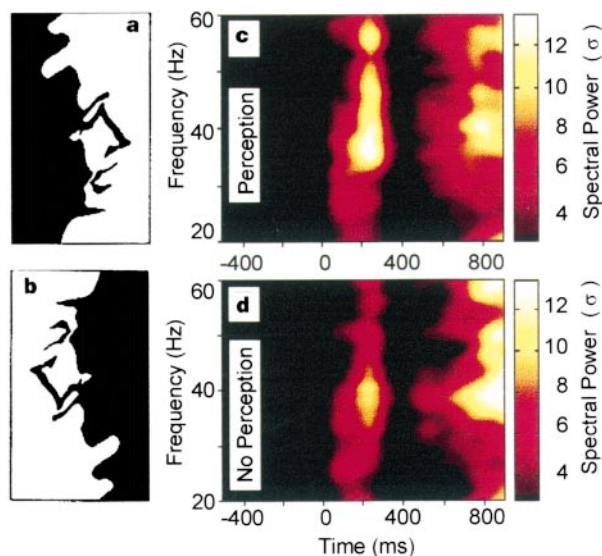


Figure 1 Stimuli and emission time–frequency charts. **a, b**, Examples of 'Mooney' faces, high-contrast pictures of a human face. These pictures are easily recognized as human faces when seen upright (**a**), but are difficult to recognize when inverted (**b**). **c, d**, Spectral power following stimulation. A time–frequency transform was computed in each trial and then summed over all trials, subjects and electrodes. The chart retains mainly the induced component of the gamma response. Both charts exhibit two periods of increased gamma-power emission (between 20 and 60 Hz). Power peaks at ~ 230 ms after stimulus onset, and between 33 and 39 Hz. The perception condition elicits a significantly stronger gamma response than the no-perception condition (Wilcoxon $T = 5$, $Z = 2.29$, $P < 0.05$). The second peak lies at ~ 800 ms and 40 ± 5 Hz; it follows after the reaction time (645 ± 20 ms for perception; 766 ± 22 ms for no-perception) and no significant differences between conditions are found. The colour scale is indicated in standard deviations, calculated from the 500-ms baseline. We are summarizing our results here, but 7 out of 10 individual results were also significant.

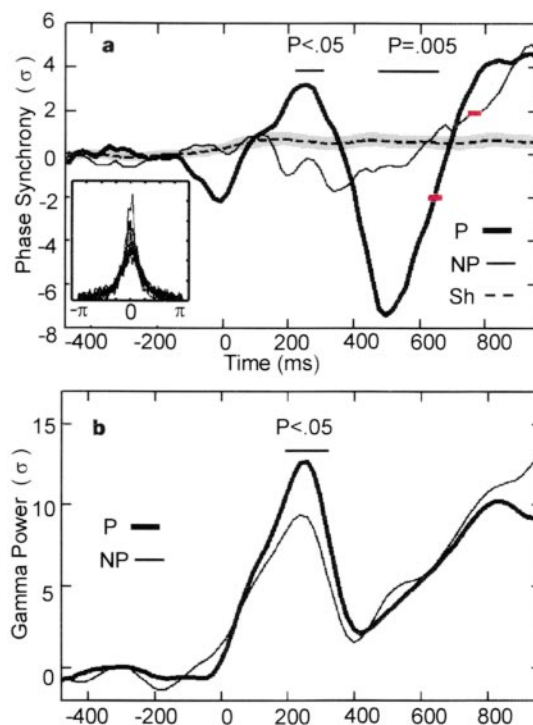


Figure 2 Time courses of phase synchrony and gamma activity. Results shown are averages over trials, electrodes, and subjects. Values are in standard deviations from the 500-ms baseline. Thick line, face-perception condition (P); thin line, no-perception condition (NP); dashed line, synchrony computed on shuffled data (Sh); the grey strip indicates dispersion of these data ± 3 standard errors; horizontal red lines indicate $\pm$ standard error of reaction time. **a**, Phase synchrony for the P and NP conditions. NP synchrony remains stable and near the shuffling average until 700 ms. Phase synchrony for the P condition increases at 230 ms ($P < 0.05$), decreases sharply at 500 ms ($P = 0.005$; absolute value of 0.3), and ends with a second increase. Synchrony is measured with a 250-ms long wavelet; a commensurable spread of synchrony follows, thus explaining results that seem to begin before stimulation onset. Inset, phase-lag distributions for ten increasingly distant synchronic electrode pairs; for all distributions $\mu = 0$ and $\sigma \approx 40^\circ$. **b**, Gamma-power activity in the 34–40 Hz band. In contrast to phase synchrony, and despite the presence of a significant difference in synchrony at 250 ms, both conditions follow a similar time course of gamma-band activity.

introduced an effective method²² with which to directly measure phase-locking; this method is based on wavelet filtering followed by trial-by-trial comparison of phase differences, and overcomes the shortcomings of other techniques used previously^{5,23,24}. Electrical activity was taken to be synchronous if phase lag between two electrodes remained constant throughout all the trials.

The main results obtained from our studies of phase synchrony are summarized in Fig. 2a and compared to gamma activity in Fig. 2b. Phase synchrony differed markedly between the perception and no-perception conditions (Fig. 2a). Only the perception condition elicited a three-part temporal pattern spanning the entire duration of the task, from stimulus presentation to motor response. First, a significant increase in synchrony for the perception relative to the no-perception condition occurred at ~200–260 ms after stimulus presentation (Wilcoxon $T = 8$, $Z = 1.99$, $P < 0.05$); the temporal coincidence of this increase in synchrony with the first gamma response indicates a functional involvement of this increase in perception itself. It was followed by a marked decrease in synchrony, or desynchronization, centred at 500 ms (Wilcoxon $T = 0$, $Z = 2.8$, $P < 0.005$). A final increase in synchrony was present for both conditions at about the reaction time (645 ± 20 ms of standard error for perception; 766 ± 22 ms for non-perception). Furthermore, synchronic electrodes had zero-centred phase-lag distributions, regardless of interelectrode distance (Fig. 2, inset). In contrast to synchrony, gamma activity for both the perception and the no-perception conditions did not yield qualitative differences, but only an amplitude difference at 230 ms (Fig. 2b).

The stage of sharp decrease of phase synchrony has not been described before, to our knowledge, but it appears as a very significant effect in our results ($P = 0.005$). Given the method used for computing synchrony, this decrease cannot be interpreted as a return to the baseline level of synchrony (see Methods). We suggest that it reflects a process of active desynchronization. This result provides the first direct support for the previous proposal³ that a transition between two distinct cognitive acts (such as face perception and motor response) should be punctuated by a transient stage of undoing the preceding synchrony and allowing for the emergence of a new ensemble, through cellular mechanisms that remain to be established.

To validate our results, we compared them with synchronies computed in shuffled trials, a technique used widely in single-cell studies^{4,9,10}. Shuffling is done by randomizing the order of trials and calculating synchronies between events that were not recorded at the same time. This allows an estimation of the magnitude of the background random fluctuations for the values of synchrony as measured here. The average of shuffled synchronies varied little over the entire observation window and showed limited variance, in contrast to both of our experimental conditions (Fig. 2a). Thus our results do not represent random phase coincidences present in scalp recordings.

Detailed spatiotemporal information is provided by the regional distribution of gamma activity and phase synchrony over the scalp (Fig. 3). The pattern of gamma activity was spatially homogeneous and was similar between the perception and no-perception conditions over time, differing only in amplitude. In contrast, the pattern of synchrony was spatially unhomogeneous and differed over time between conditions. Compared with the no-perception condition, which showed few synchronous patterns, the perception condition exhibited a sequence of localized spatial patterns that evolved over time (Fig. 3). Synchrony first increased in the area between the left parieto-occipital and frontotemporal regions. Desynchronization was then observed between the parietal and occipitotemporal areas bilaterally. Parietal regions are involved in visual perception^{25,26} and episodic memory²⁷. Interactions between these regions and occipitotemporal regions, in particular the fusiform gyrus, have been linked to perceptual learning of degraded

Mooney-like faces²⁵. We propose that phase interactions between parietal and occipitotemporal regions are essential in the large-scale integration that is needed for the perception of upright Mooneys as faces. The second synchrony increase, which is probably linked to motor response, was predominant between the right temporal and central regions. Only during this second period of increase were there slight similarities between the perception and no-perception conditions, because the subject responded in both cases.

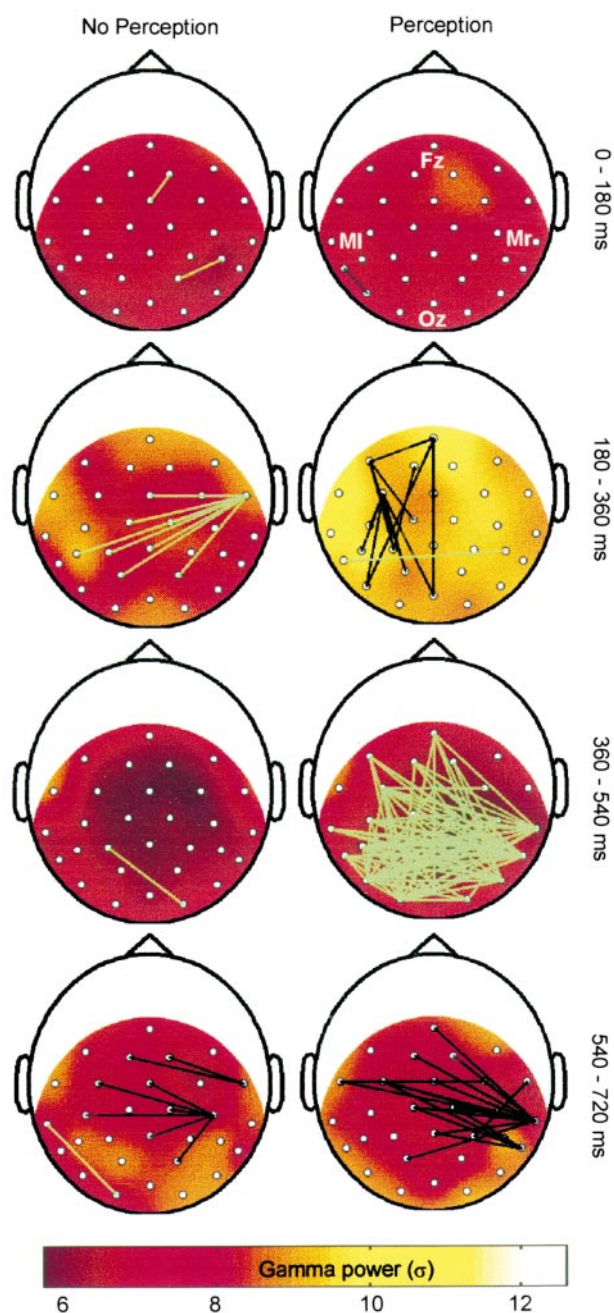


Figure 3 The shadow of a perception. Average scalp distribution of gamma activity and phase synchrony. Colour coding indicates gamma power (averaged in a 34–40-Hz frequency range) over an electrode and during a 180-ms time window, from stimulation onset (0 ms) to motor response (720 ms). Gamma activity is spatially homogeneous and similar between conditions over time. In contrast, phase synchrony is markedly regional and differs between conditions. Synchrony between electrode pairs is indicated by lines, which are drawn only if the synchrony value is beyond the distribution of shuffled data sets ($P < 0.01$; see Methods). Black and green lines correspond to a significant increase or decrease in synchrony, respectively.

The biological significance of phase synchrony computed from scalp EEGs has been doubted because it is difficult to rule out the occurrence of spurious synchronization resulting from volume conduction²³. However, such spurious synchronization cannot account for our results. Volume conduction induces a pattern of synchronization that decays rapidly as the separation between electrodes increases beyond 2 cm on the surface of the cortex²³. In contrast, we found that synchrony can be established between recording sites situated far away from each other on the head surface; only 7% of the synchronies reported here were between neighbouring recording sites. It is also conceivable that distant synchronization could result from a powerful deep source that diffuses widely over the scalp. But if this were the case, the phase-synchrony pattern should coincide with gamma activity, the electrodes with higher emission being the most synchronous ones. Our results do not show this (Fig. 3). Finally, the desynchronizing periods found here are impossible to explain by volume conduction. If synchrony effects were just a reflection of gamma activity then desynchronization should be associated with periods of low gamma activity. Our results show the opposite effect: desynchronization co-existed with periods of above-average gamma activity. Furthermore, the same gamma level led to strong desynchronization in the perception condition but had no effect on the no-perception condition (Fig. 2). Finally, the zero-centred phase-lag distribution found here also suggests that the measured synchronies did not have an artefactual origin.

The finding of gamma oscillations has often been taken, erroneously, as an indication of synchrony. Indeed, changes in gamma-band spectral content cannot be conflated with the phase synchrony between pairs of electrodes at which such gamma activity occurs. Only synchrony measures bear directly on the possible role of gamma activity in cognition, as synchrony provides direct information about electrode pairs and their regional location, which power emission alone cannot provide. To our knowledge, our results are the first to support the theory that phase synchrony is directly involved in human cognition. The long-range character of the phase synchrony indicates that gamma-phase synchrony (and desynchrony) may be viewed as a mechanism that subserves large-scale cognitive integration^{2,3,5,8}, and not just local visual-feature binding. Finally, we stress that the detection of phase synchrony/desynchrony over the scalp amounts to a dynamic brain mapping that is essential for the study of the neural basis of cognitive tasks. □

Methods

Protocol and recordings. Experimental design¹⁹ and data analysis²² have been described in detail elsewhere. In brief, Mooney faces were presented randomly for 200 ms in upright or inverted position, in a total of 320 trials. Ten subjects (20–30 years old; seven females) gave their response by pressing buttons situated under their right and left index fingers. EEGs were recorded by 30 electrodes at standard 10–20 positions, to which we added a lower row (M1, M2, P9, P10, PO9, PO10, O9, and O10). Sampling was taken at 500 Hz.

Time–frequency analysis. After automatic correction of eye-movement artefact trials, signals were high-pass-filtered at 15 Hz and time–frequency-analysed using the pseudo Wigner–Ville transformation^{17,20}. Resulting time–frequency maps were normalized and averaged through trials, electrodes and subjects.

Phase-synchrony detection. Previous methods for measuring phase synchrony between electrode pairs have included spectral coherence^{5,23}, which mixes energy and phase information, and detection of maximal values after filtering²⁴, which is inaccurate and slow when large data sets are involved. In the method introduced here, for each subject phase synchrony was computed only for the frequency f_0 of his/her maximal gamma activity (varying from 35 to 45 Hz, depending on the subject). Phase was measured from narrow-band-filtered signals ($f_0 \pm 3$ Hz) by convolution with a complex Gabor wavelet designed for f_0 . An instantaneous phase value, $\varphi_i(f_0, t, k)$, which is a complex number of unit magnitude, was thus obtained for one electrode, i , a chosen frequency, f_0 , at time bin indexed by t , and trial k . For each electrode pair, i and j , and time t , and for all of the $k = 1, \dots, N$ trials, a global

phase-locking value $\varphi_{ij}(f_0, t)$ is computed as:

$$\varphi_{ij}(f, t) = \frac{\left| \sum_k \varphi_i - \varphi_j \right|}{N}$$

φ_{ij} is a real value bounded between 1 (if phase difference is constant) and 0 (if phase difference is random).

Normalization. To calculate synchrony values comparable between near (<2 cm) and distant electrode pairs, we carried out a normalization procedure so that the $\varphi_{ij}(f, t)$ values were compared with the 500-ms baseline preceding the stimulus. Given φ_{ij} , let μ_{ij} and σ_{ij} be the mean and standard deviation computed from a 500-ms prestimulus baseline; the normalized phase-locking values are then computed as $\tilde{\varphi}_{ij} = (\varphi_{ij} - \mu_{ij})/\sigma_{ij}$. The same normalization procedure is applied to time–frequency matrices on a frequency-by-frequency basis.

Topographical synchrony. To display the lines indicating synchrony over individual pairs of electrodes (Fig. 3) we used the following statistical procedure. Let W_k be a 180-ms time window between stimulus arrival and motor response, and let $\varphi_{ij}(W_k)$ be the average phase synchrony between electrodes i and j over the entire time window, W_k . To enhance time changes, we compared W_k with the previous time window, W_{k-1} , and defined the phase-locking value between pairs finally as $\Delta\varphi_{ij}(W_k) = \varphi_{ij}(W_k) - \varphi_{ij}(W_{k-1})$. For each $\Delta\varphi_{ij}(W_k)$, 200 values were analogously computed on shuffled data $\Delta\varphi_{ij}^s(W_k)$. A $\Delta\varphi_{ij}$ value is retained as statistically significant only if greater than (or lesser than) any of the 200 shuffled values $\Delta\varphi_{ij}^s$, thus corresponding to a two-tailed probability value of $P = 0.01$.

Received 7 October; accepted 23 December 1998.

- Freeman, W. J. *Mass Action in the Nervous System* (Academic, New York, 1975).
- Damasio, A. R. Synchronous activation in multiple cortical regions: a mechanism for recall. *Semin. Neurosci.* **2**, 287–297 (1990).
- Varela, F. J. Resonant cell assemblies: a new approach to cognitive function and neuronal synchrony. *Biol. Res.* **28**, 81–95 (1995).
- Singer, W. & Gray, C. M. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* **18**, 555–586 (1995).
- Bressler, S. L., Coppola, R. & Nakamura, R. Episodic multiregional cortical coherence at multiple frequencies during visual task performance. *Nature* **366**, 153–156 (1993).
- Gray, C. M., König, P., Engel, A. K. & Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–337 (1989).
- Vaadia, E. *et al.* Dynamics of neuronal interaction in the monkey cortex in relation to behavioral events. *Nature* **373**, 515–518 (1995).
- Roelfsema, P. R., Engel, A. K., König, P. & Singer, W. Visuomotor integration is associated with the zero time-lag synchronization among cortical areas. *Nature* **385**, 157–161 (1997).
- Neuenschwander, S., Engel, A., König, P., Singer, W. & Varela, F. J. Synchronization of neuronal responses in the optic tectum of awake pigeons. *Vis. Neurosci.* **13**, 575–584 (1996).
- Kreiter, A. K. & Singer, W. Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey. *J. Neurosci.* **16**, 2381–2396 (1996).
- Llinas, R. & Ribary, U. Coherent 40 Hz oscillation characterizes dream state in humans. *Proc. Natl Acad. Sci. USA* **90**, 2078–2081 (1993).
- Tiitinen, H. *et al.* Selective attention enhances the auditory 40-Hz transient response in humans. *Nature* **364**, 59–60 (1993).
- Desmedt, J. E. & Tomberg, C. Transient phase-locking of 40 Hz electrical oscillations in prefrontal parietal cortex reflects the process of conscious somatic perception. *Neurosci. Lett.* **168**, 126–129 (1994).
- Joliot, M., Ribary, U. & Llinas, R. Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. *Proc. Natl Acad. Sci. USA* **91**, 11748–11751 (1994).
- Pantev, C. Evoked and induced gamma band activity of the human cortex. *Brain Topogr.* **7**, 321–330 (1995).
- Tallon-Baudry, C., Bertrand, O., Delpuech, C. & Pernier, J. Oscillatory gamma band (30–70 Hz) activity induced by a visual search task in human. *J. Neurosci.* **17**, 722–734 (1997).
- Lachaux, J. P. *et al.* Gamma band activity in human intracortical recordings triggered by cognitive tasks. *Eur. J. Neurosci.* (submitted).
- Mooney, C. M. Age in the development of closure ability in children. *Can. J. Psychol.* **11**, 219–226 (1957).
- George, N., Jemel, B., Fiori, N. & Renault, B. Face and shape repetition effects in humans: a spatio-temporal ERP study. *NeuroReport* **8**, 1417–1423 (1997).
- Mecklenbräuer, W. F. G. (ed.) *The Wigner Distribution—Theory and Applications in Signal Processing* (Elsevier, Amsterdam, 1993).
- Jeffreys, D. A. Evoked potentials studies of face and object processing. *Vis. Cogn.* **3**, 1–38 (1996).
- Lachaux, J. P., Rodriguez, E., Müller-Gerking, J., Martinier, J. & Varela, F. J. Measuring phase-synchrony in brain signals. *Hum. Brain Mapp.* (submitted).
- Menon, V. *et al.* Spatio-temporal correlations in human gamma band electrocorticograms. *Electroencephalogr. Clin. Neurophysiol.* **98**, 89–102 (1996).
- Yordanova, J., Kolev, V. & Demiralp, T. The phase-locking of auditory gamma band responses in humans is sensitive to task processing. *NeuroReport* **8**, 3999–4004 (1997).
- Dolan, R. J. *et al.* How the brain learns to see objects and faces in an impoverished context. *Nature* **389**, 596–599 (1997).
- Friedman-Hill, S. R., Robertson, L. C. & Treisman, A. Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. *Science* **269**, 853–855 (1995).
- Shallice, T. *et al.* Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature* **368**, 633–635 (1994).

Acknowledgements. We thank Y. Okada for suggestions on the manuscript. This work was supported by grants from MIDEPLAN (Chile), DRET (France), and the Human Science Frontier.

Correspondence and requests for materials should be addressed to E.J.V. (e-mail: fv@ccr.jussieu.fr).

Coherence of gamma-band EEG activity as a basis for associative learning

Wolfgang H. R. Miltner*, Christoph Braun†, Matthias Arnold‡, Herbert Witte‡ & Edward Taub§

* Department of Biological and Clinical Psychology, Institute of Psychology, Friedrich-Schiller-University, Am Steiger 3/1, D-07743 Jena, Germany

† Institute of Medical Psychology and Behavioral Neuroscience, Eberhard-Kals-University, Hoppe-Seyler-Str. 3, D-72076 Tübingen, Germany

‡ Institute of Medical Statistics, Informatics, and Documentation, Friedrich-Schiller-University, Jahnstraße 1, D-07743 Jena, Germany

§ Department of Psychology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

Different regions of the brain must communicate with each other to provide the basis for the integration of sensory information, sensory-motor coordination and many other functions that are critical for learning, memory, information processing, perception and the behaviour of organisms. Hebb¹ suggested that this is accomplished by the formation of assemblies of cells whose synaptic linkages are strengthened whenever the cells are activated or 'ignited' synchronously. Hebb's seminal concept has intrigued investigators since its formulation, but the technology to demonstrate its existence had been lacking until the past decade. Previous studies have shown that very fast electroencephalographic activity in the gamma band (20–70 Hz) increases during, and may be involved in, the formation of percepts and memory^{2–6}, linguistic processing⁷, and other behavioural and preceptual functions^{8–12}. We show here that increased gamma-band activity is also involved in associative learning. In addition, we find that another measure, gamma-band coherence, increases between regions of the brain that receive the two classes of stimuli involved in an associative-learning procedure in humans. An increase in coherence could fulfil the criteria required for the formation of hebbian cell assemblies¹, binding together parts of the brain that must communicate with one another in order for associative learning to take place. In this way, coherence may be a signature for this and other types of learning.

The subjects were 16 (9 female) healthy, right-handed (Edinburgh test of handedness¹³) student volunteers, aged between 22 and 36 years. Before participation, subjects were informed about all aspects of the experiment in accordance with the Helsinki Convention on human experimentation, and all signed an informed consent. The experimental protocol was approved by the institution's ethics committee.

In the associative-learning phase of the experiment, there were 60 trials in which a room-illuminating light of 3-s duration and of one colour (CS⁺), either red or green, was accompanied at its termination by an electric shock (an unconditioned stimulus (UCS)) delivered to either the dominant (8 subjects) or the non-dominant (8 subjects) mid-volar tip of the third digit. These trials were interspersed randomly with 60 trials in which a light of the alternate colour was presented for 3 s without noxious stimulation (CS[−]). The colour of CS⁺ was counterbalanced across subjects. The intertrial interval was 4 s. After the 120 acquisition trials, an extinction series of 80 trials was given in which 40 CS⁺ stimuli without electric shock and 40 CS[−] stimuli were given in random order.

Figure 1 shows the sites of the scalp electrodes from which recordings were made. The abundance of gamma activity increased at all occipital and parietal electrode sites in CS⁺ and CS[−] for the first 2 s of the trial period, after which it stayed at approximately the same elevated level for the last second of the trial ($F(5, 55) = 10.54$, $P < 0.003$). For the first 2 s of the trial there was a trend for gamma

abundance to be significantly greater in CS⁺ than in CS[−] ($0.06 > P > 0.05$).

Arrows connect the 13 pairs of electrode sites for which an analysis of variance showed that gamma coherence in the 37–43-Hz band was significantly greater during CS⁺ trials than during CS[−] trials ($P \leq 0.06$) for the 250-ms time window just before UCS onset (Fig. 1). When shock was delivered to the non-dominant (left) hand (Fig. 1a), four electrode pairs spanning the primary and association visual cortices, which receive visual input concerning CS⁺ and CS[−], and the contralateral (right) or midline pericentral cortex, which receives input from the finger to which the electric shock was applied, exhibited greater gamma coherence during CS⁺ than during CS[−], as did two electrode pairs connecting visual cortex with the contralateral posterior parietal somatosensory association area. No electrode pairs connecting visual cortex and the ipsilateral (left) pericentral cortex showed increased coherence. When shock was delivered to the dominant (right) hand (Fig. 1b), the laterality of increased coherence reversed. Five electrode pairs spanning the visual and left or midline pericentral cortex showed significantly increased coherence, as did two electrode pairs spanning visual and left posterior parietal cortices, whereas no electrode pairs connecting visual cortex and right pericentral cortex showed increased coherence. Thus, the pericentral and posterior parietal electrode sites showing a significant effect were either midline or in the hemisphere representing the finger receiving the electric shock. This temporal and stimulus specificity in combination with the topographic specificity exhibited with respect to the laterality of the UCS presentation indicates that a general attention and/or arousal effect is unlikely to produce these results, as noted in a previous review¹⁴. In the frequencies above and below the 37–43-Hz band there was an abrupt drop-off of electrode pairs exhibiting significant coherence, with approximately half or less than half the number found over the range 37–43-Hz. In the 30–37-Hz range, seven electrode pairs showed significant coherence, three when the shock was delivered to the left hand (one extending anomalously from occipital cortex to the left side of the brain) and four when the shock was delivered to the right hand (one anomalous). In the 43–48-Hz range, five electrode pairs showed significantly increased coherence, two with shock to the left hand (one anomalous) and three with shock to the right hand. Thus, the main focus of coherent gamma activity was in the 37–43-Hz band.

Gamma coherence was not greater in CS⁺ than CS[−] during, first, the 250-ms time window early in the trial (1,250–1,500 ms after CS

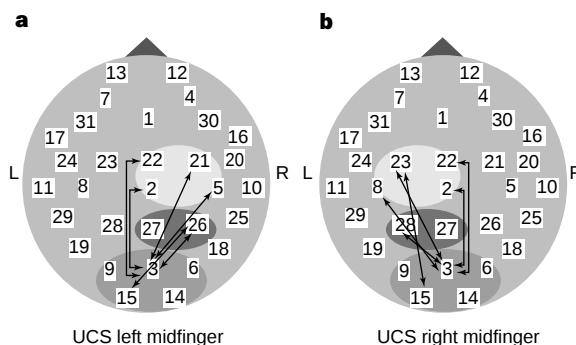


Figure 1 Pairs of electrode sites displaying significantly greater coherence in the 37–43-Hz gamma band during CS⁺ than during CS[−] trials in the 2,750–3,000-ms time window for the 120 acquisition trials. Data are pooled for all subjects in a group. **a**, Arrows connect the electrode pairs showing this effect for a shock delivered to the non-dominant hand; **b**, arrows show the gamma-coherence connections for shock delivery to the dominant hand. The differently shaded ovals denote the approximate locations of the visual cortex (rear; intermediately shaded ovals) and of the two somatosensory association (middle, darkly shaded ovals) and two primary somatosensory (front; lightly shaded ovals) cortices representing the (left or right) hand receiving the electric shock.

onset); second, the last five trials of the 80-trial extinction period (during which the shock was never presented); or third, the last 30 CS⁺ and 30 CS⁻ trials of the extinction period for the pairs of electrodes (one left and one right) covering the somatosensory/visual cortex site (3–5 and 3–8, respectively) which showed particularly strong gamma-coherence conditioning effects. Gamma coherence therefore showed a specificity of effect to that part of the trial period and training/extinction series when the UCS was presented.

Increased coherence during the trial period was limited for other frequency bands in the electroencephalographic (EEG) spectrum. There was no greater coherence in CS⁺ than in CS⁻ for any electrode pair for the delta, theta or beta bands. For the alpha 1 band, two electrode pairs showed greater coherence in CS⁺ than in CS⁻; both extended from visual cortex to somatosensory association (but not primary somatosensory) cortex, one to the side representing the finger receiving the UCS and the other to the side representing the unstimulated finger. The latter, anomalous linkage was the only one that showed a significant effect for the alpha 2 band. There was no significant increase in coherence between any electrode pairs in the 80–100-Hz range ($F(1, 15) = 0.21$, not significant), indicating that the results were not due to an artefact of 'spillover' muscle activity.

Our observation of increased gamma coherence during the time window preceding the application of the UCS in CS⁺ trials was specific to the application of time-varying autoregressive modelling of the multivariate time series, which is based on Kalman filtering¹⁵ where coherence measures are estimated adaptively. When we applied conventional coherence methods using cross-spectral analysis based on the fast Fourier transform (FFT) or wavelet analysis, similar results could not be obtained. Presumably, the

reason is that FFT is not sufficiently effective in quantifying dynamic signals as it requires a signal to be stationary during analysis of the whole interval.

The contingent negative variation (CNV), which is generally considered representative of an increase in neuronal excitability in neural networks in preparation for carrying out a task, and which is a useful index of learning¹⁶, was greater for CS⁺ than CS⁻ in each 500-ms time period after trial onset, and was significant at the somatosensory and motor electrode sites in the last two 500-ms time periods before shock onset ($F(1, 15) = 8.43$ and 6.59 , $P < 0.01$ and $P < 0.02$, respectively). The difference in evoked-response-potential waves between CS⁺ and CS⁻ (CS⁺ minus CS⁻) at each pericentral and posterior parietal site exhibiting increased gamma coherence with an occipital site (receiving the CSs) is shown in Fig. 2. The height of the upward (negative) deflection of these waves indicates the substantial extent to which the CNV to CS⁺ exceeded the response to CS⁻ (which was essentially zero at all electrode sites). The change in CNV resulting from the training supplements the change in colour-preference data (see Methods) as indicators, independent from the gamma-band data, that learning had taken place.

Our study, which is based on a previous classical conditioning/CNV study¹⁷, extends to associative learning, the process of greatest interest to Hebb, the phenomena to which gamma-band EEG activity may be related. Moreover, the formation of stable connections between different brain areas involved in the processing of stimuli presented to different sensory modalities during an associative-learning procedure is a good model of the formation of cell assemblies, as it provides an unambiguous and easily specified means of requiring that widely separated areas of the brain interact. It is of further interest that not only did gamma-band abundance increase in regions of the brain that received the two classes of stimuli involved in the training procedure, but gamma-band coherence also increased between them. Coherence or in-phase synchronicity is, in conceptual terms, the simplest and most straightforward process that could provide the basis for the formation of hebbian cell assemblies. The linkage or communication between visual cortex and somatosensory/motor cortex that was observed could have been direct, mediated by a third, intermediary region(s), or produced by the oscillation in a third region(s) that caused the gamma coherence in both areas. At the neuronal level, the mechanism could involve the recently discovered superficial pyramidal 'chattering cells' that may act as excitatory pacemakers for coherent gamma-band rhythms^{18,19} and/or the interconnected chaining of interneuronal networks operating to produce 'spike doublets'^{20,21}. It remains for future research to resolve these issues. Whatever the mechanism, however, it seems possible that the development of gamma coherence is involved not only in the establishment of associative learning, but also in other types of learning as well. □

Methods

The experiment was carried out in a sound-retarded, electrically shielded chamber. The subjects sat in a reclining chair and were asked to keep their arms relaxed on arm rests and their eyes open. They were paid for the test session, which took ~3 h.

Training. Intracutaneous electrical stimuli²² (10-ms duration) were used as painful UCS. Pain threshold was determined before the training phase by the method of limits using three pairs of ascending and descending series of stimulus intensities. Pain intensity was judged on a scale from 1 (not painful at all) to 12 (unbearably painful). The mean stimulation current of all trials for pain-threshold determination with an intensity rating of 5 was adopted as the pain threshold. Electrical stimulation intensity was set 40% above the physical value of the pain threshold.

EEG recording. EEG data were recorded from 31 Ag/AgCl electrodes mounted on the subject's scalp according to the international 10–20 system, with additional electrodes interspersed between standard electrode sites, and at left

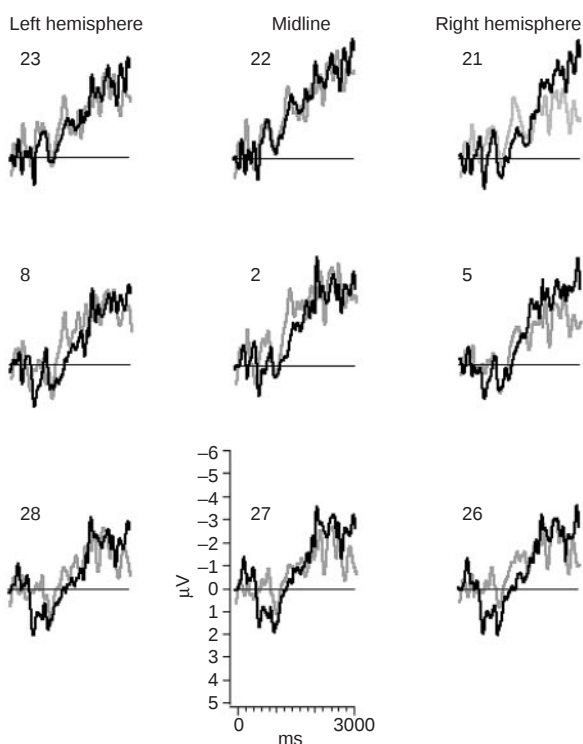


Figure 2 Difference in evoked-response-potential waves between CS⁺ and CS⁻ at each electrode involved in response to the UCS when the UCS was administered to the left (grey line) and right (black line) midfingers. A CNV (upward deflection of wave) developed at each of these electrode sites in the period before UCS presentation during CS⁺ but not CS⁻ (repeated-measures analysis of variance, all P values < 0.05). The CNV to the CS⁺ indicates readiness to respond to the UCS and that learning had taken place. The number in the top left corner of each panel indicates for what electrode in Fig. 1 the CNV waveforms are shown.

and right earlobes, as outlined in Fig. 1. Electrode Cz served as common reference. Vertical and horizontal electro-oculographic (EOG) data were recorded for eye blinks and eye movements. EEG and EOG data were collected for a total period of 4,250 ms, beginning 250 ms before visual-stimulus presentation and continuing for 1,000 ms after the end of a trial. EEG data were corrected for eye movements and eye blinks²³. Data were sampled online at 200 Hz using a time constant of 10 s.

Data analysis. For off-line data analysis, a current-source density (CSD) analysis was performed based on a spherical spline interpolation algorithm using two dimensional Laplace-operators and weighted Legendre polynomials^{24–26} in order to maximize the topographic specificity of electrical sources and sinks and to obtain reference-free measures for each electrode²⁴. Resulting CSD waveforms were then submitted to a method of coherence analysis involving time-varying autoregressive modelling of multivariate time series based on Kalman filtering^{15,27–28}. In the case of single-component signals the use of Kalman filters to fit autoregressive models with time-varying coefficients is common^{27–28}. This is possible because such signal models can be given in state-space form. As the Kalman filter can be designed for multi-component signals, we have used a similar state-space form for multi-component autoregressive models with time-varying coefficients. This enables an adaptive estimation of the autoregressive coefficients and derived spectral parameters (that is, coherence) which is suitable for the analysis of non-stationary signals. An order of 16 was chosen for the fitted autoregressive models.

Coherence was computed separately in the bandwidth 37–43 Hz (refs 2, 19), for the delta (0.6–3.5 Hz), theta (3.6–7.5 Hz), alpha 1 (7.6–10 Hz) and alpha 2 (10.1–12.5 Hz) frequency bands, and for additional bands between 30 and 36 Hz, 43 and 48 Hz, and 80 to 100 Hz for the 250-ms time window from 1,250 to 1,500 ms after stimulus onset and the window just before shock onset (that is, 2,750–3,000 ms after stimulus onset). This was also done for the raw EEG. The pairs of electrodes for which analysis was done covered the occipital area (CS⁺, CS⁻), the primary and association somatosensory projection areas (noxious stimulus), and the primary motor areas on both sides of the brain and at the midline. Gamma abundance was also calculated for the frequency band studied in refs 2, 6 (37–43 Hz). Coherence measures were transformed using Fisher's z-transformation and collapsed to obtain mean coherence measures for each subject, each pair of electrodes and each condition of visual stimulation.

Preference for colours of CS⁺ and CS⁻. Data on colour preference were obtained by questionnaire for the two colours (red and green) assigned randomly to CS⁺ and CS⁻ before, during, and at the end of acquisition and during and at the end of extinction. The colour-preference data were submitted to a repeated-measures analysis of variance and analysed as a function of condition (CS⁺, CS⁻), and time (beginning, during and end of acquisition, and during and end of extinction). Results were corrected for violation of sphericity using the Greenhouse–Geisser approach to epsilon correction of degrees of freedom. There was a significant effect for condition ($F(1, 18) = 6.64$, $P < 0.02$), time ($F(4, 72) = 10.43$, $P < 0.0001$, $e = 0.60$), and the condition $\times$ time interaction ($F(4, 72) = 9.14$, $P < 0.0001$, $e = 0.751$), indicating that learning of the association between colour and electric shock had taken place. CS⁺ was perceived as most aversive during training, whereas during extinction the difference in preference for CS⁺ over CS⁻ had almost disappeared.

Received 12 October; accepted 1 December 1998.

- Hebb, D. O. *The Organization of Behavior* (Wiley, New York, 1949).
- Singer, W. & Gray, C. M. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* **18**, 555–586 (1995).
- Basar, E., Basar-Eroglu, C. & Schürmann, M. Sensory and cognitive components of brain resonance responses. *Acta Otolaryngol. (Stockh.)* **491**, 25–35 (1991).
- Engel, A. K., König, P., Kreiter, A. K. & Singer, W. Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. *Science* **252**, 1177–1179 (1991).
- Pantev, C. Evoked and induced gamma-band activity of the human cortex. *Brain Topogr.* **7**, 321–330 (1995).
- Singer, W. The formation of cooperative cell assemblies in the visual cortex. *J. Exp. Biol.* **153**, 177–197 (1990).
- Pulvermüller, F., Lutzenberger, W., Preißl, H. & Birbaumer, N. Spectral responses in the gamma-band: physiological signs of higher cognitive processes? *Neuroreport* **6**, 2059–2064 (1995).
- Murthy, V. N., Aoki, F. & Fetz, E. E. In *Oscillatory Event-Related Brain Dynamics* (eds Pantev, C., Elbert, T. & Lütkenhöner, B.) 213–226 (Plenum, New York, 1994).
- Llinas, R. R. & Ribary, U. Coherent 40-Hz oscillation characterizes dream state in humans. *Proc. Natl. Acad. Sci. USA* **90**, 2078–2081 (1993).
- Sheer, D. E. In *Self-regulation of the Brain and Behavior* (eds Elbert, T., Rockstroh, B., Lutzenberger, W. & Birbaumer, N.) 64–84 (Springer, Berlin, 1984).
- Steriade, M., Amzica, F. & Contreras, D. Synchronization of fast (30–40 Hz) spontaneous cortical rhythms during brain activation. *J. Neurosci.* **16**, 392–417 (1996).

12. Tiitinen, H. et al. Selective attention enhances the auditory 40-Hz transient response in humans. *Nature* **364**, 59–60 (1993).
13. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychology* **9**, 97–113 (1971).
14. Pulvermüller, F., Birbaumer, N., Lutzenberger, W. & Mohr, B. High-frequency brain activity: its possible role in attention, perception and language processing. *Prog. Neurobiol.* **52**, 427–445 (1997).
15. Arnold, M., Miltner, W. H. R., Bauer, R. & Braun, C. Adaptive AR modeling of nonstationary time series by means of Kalman filtering. *IEEE Trans. Biomed. Eng.* **45**, 553–562 (1998).
16. Rockstroh, B., Elbert, T., Canavan, A., Lutzenberger, W. & Birbaumer, N. *Slow Cortical Potentials and Behaviour* (Urban & Schwarzenberg, Munich, 1989).
17. Waschulewski Floruss, H., Miltner, W., Brody, S. & Braun, C. Classical conditioning of pain responses. *Int. J. Neurosci.* **78**, 21–32 (1994).
18. Gray, C. M. & McCormick, D. A. Chattering cells: superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex. *Science* **274**, 109–113 (1996).
19. Gray, C. M., König, P., Engel, A. K. & Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–337 (1989).
20. Jefferys, J. G., Traub, R. D. & Whittington, M. A. Neuronal networks for induced '40 Hz' rhythms. *Trends Neurosci.* **19**, 202–208 (1996).
21. Traub, R. D., Whittington, M. A., Stanford, I. M. & Jefferys, J. G. A mechanism for generation of long-range synchronous fast oscillations in the cortex. *Nature* **383**, 621–624 (1996).
22. Bromm, B. & Meier, W. The intracutaneous stimulus: a new pain model for algometric studies. *Meth. Find. Exp. Clin. Pharmacol.* **6**, 405–410 (1984).
23. Gratton, G., Coles, M. G. & Donchin, E. A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* **55**, 468–484 (1983).
24. Perrin, F., Pernier, J., Bertrand, O. & Echallier, J. F. Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* **72**, 184–187 (1989).
25. Biggins, C. A., Fein, G., Raz, J. & Amir, A. Artificially high coherences result from using spherical spline computation of scalp current density. *Electroencephalogr. Clin. Neurophysiol.* **79**, 413–419 (1991).
26. Perrin, F. Comments on article by Biggins et al. *Electroencephalogr. Clin. Neurophysiol.* **83**, 171–174 (1992).
27. Haykin, S. *Adaptive Filter Theory* (Prentice-Hall, Englewood Cliffs, New Jersey, 1986).
28. Chen, H. H. & Guo, L. *Identification and Stochastic Adaptive Control* (Birkhäuser, Boston, Massachusetts, 1991).

Acknowledgements. We thank I. Gutberlet for assistance in data analysis. This research was supported by a grant from the Deutsche Forschungsgemeinschaft to W.H.R.M. and a grant from the Rehabilitation Research and Development service, US Department of Veterans Affairs to E.T.

Correspondence and requests for materials should be addressed to W.H.R.M. (email: miltner@biopsy.uni-jena.de).

Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*

Feng Gao*, Elizabeth Bailes†, David L. Robertson‡, Yalu Chen*, Cynthia M. Rodenburg*, Scott F. Michael*§, Larry B. Cummins||, Larry O. Arthur†, Martine Peeters#, George M. Shaw*☆, Paul M. Sharp† & Beatrice H. Hahn*

* Departments of Medicine and Microbiology, University of Alabama at Birmingham, 701 S. 19th Street, LHRB 613, Birmingham, Alabama 35294, USA

† Institute of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, UK

‡ Laboratory of Structural and Genetic Information, CNRS, Marseilles 13402, France

§ Southwest Foundation for Biomedical Research, San Antonio, Texas 78245, USA

¶ AIDS Vaccine Program, National Cancer Institute-Frederick Cancer Research and Development Center, SAIC Frederick, Frederick, Maryland 21702, USA

Laboratoire Retrovirus, ORSTOM, BP 5045, Montpellier 34032, France

☆ Howard Hughes Medical Institute, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

The human AIDS viruses human immunodeficiency virus type 1 (HIV-1) and type 2 (HIV-2) represent cross-species (zoonotic) infections^{1–4}. Although the primate reservoir of HIV-2 has been clearly identified as the sooty mangabey (*Cercocebus atys*)^{2,4–7}, the origin of HIV-1 remains uncertain. Viruses related to HIV-1 have been isolated from the common chimpanzee (*Pan troglodytes*)^{8,9}, but only three such SIVcpz infections have been documented^{1,10,11}, one of which involved a virus so divergent¹¹ that it might represent a different primate lentiviral lineage. In a search for the HIV-1 reservoir, we have now sequenced the genome of a new SIVcpz

§ Present address: Department of Tropical Medicine, Tulane University, New Orleans, Louisiana 70112, USA.

strain (SIVcpzUS) and have determined, by mitochondrial DNA analysis, the subspecies identity of all known SIVcpz-infected chimpanzees. We find that two chimpanzee subspecies in Africa, the central *P. t. troglodytes* and the eastern *P. t. schweinfurthii*, harbour SIVcpz and that their respective viruses form two highly divergent (but subspecies-specific) phylogenetic lineages. All HIV-1 strains known to infect man, including HIV-1 groups M, N and O, are closely related to just one of these SIVcpz lineages, that found in *P. t. troglodytes*. Moreover, we find that HIV-1 group N is a mosaic of SIVcpzUS- and HIV-1-related sequences, indicating an ancestral recombination event in a chimpanzee host. These results, together with the observation that the natural range of *P. t. troglodytes* coincides uniquely with areas of HIV-1 group M, N and O endemicity, indicate that *P. t. troglodytes* is the primary reservoir for HIV-1 and has been the source of at least three independent introductions of SIVcpz into the human population.

Five lines of evidence have been used to substantiate zoonotic transmission of primate lentiviruses³: first, similarities in viral genome organization; second, phylogenetic relatedness; third, prevalence in the natural host; fourth, geographic coincidence; and fifth, plausible routes of transmission. For HIV-2, a virus (SIVsm) that is genomically indistinguishable and closely related phylogenetically was found in substantial numbers of wild-living sooty mangabeys whose natural habitat coincides with the epicentre of the HIV-2 epidemic^{2,4-7}. Close contact between sooty mangabeys and humans is common because these monkeys are hunted for food and kept as pets^{6,7}. No fewer than six independent transmissions of SIVsm to humans have been proposed^{4,6,7}. In contrast, the origin of HIV-1 is much less certain³. HIV-1 is most similar in sequence and genomic organization to viruses found in chimpanzees (SIVcpz)^{1,10,11}, but a wide spectrum of diversity between HIV-1 and SIVcpz¹¹, an apparent low prevalence of SIVcpz infection in wild-living animals^{8,9,12}, and the presence of chimpanzees in geographic regions of Africa¹³ where AIDS was not initially recognized have cast doubt on chimpanzees as a natural host and reservoir for HIV-1. Rather, it has been suggested that another, as yet unidentified, primate species could be the natural host for SIVcpz and HIV-1 (refs 1, 11).

We recently identified a fourth chimpanzee with natural SIVcpz infection. This animal (Marilyn) was wild-caught in Africa (country of origin unknown), exported to the United States as an infant, and used as a breeding female in a primate facility until her death at age 26 years¹². During a serosurvey in 1985, Marilyn was the only chimpanzee of 98 tested who had antibodies strongly reactive with HIV-1 by enzyme-linked immunosorbent assay (ELISA) and western immunoblot¹². She had never been used in AIDS research and had not received human blood products after 1969. She died in 1985 after giving birth to still-born twins. An autopsy revealed endometritis, retained placental elements and sepsis as the final cause of death. Depletion of lymphoid tissues was not noted. Here we used the polymerase chain reaction (PCR) to amplify HIV- or SIV-related DNA sequences directly from uncultured (frozen) spleen and lymph-node tissue obtained at autopsy in order to characterize the infection responsible for Marilyn's HIV-1 seropositivity. Amplification and sequence analysis of subgenomic *gag* (508 base pairs (bp)) and *pol* (766 bp) fragments revealed the presence of a virus related to, but distinct from, known SIVcpz and HIV-1 strains. Because virus isolation from the autopsy tissues was unsuccessful, we used PCR to amplify and sequence four overlapping subgenomic fragments that together comprised a complete proviral genome, which we termed SIVcpzUS. Analysis of potential coding regions revealed the presence of a *vpu* gene (found only in HIV-1 and SIVcpz viruses)^{1,11,14} in addition to structural and regulatory genes common to all primate lentiviruses. None of the genes in SIVcpzUS contained deletions, insertions or rearrangements, and all reading frames were open except for *gag* (p24) and *rev* (second exon), which contained single in-frame stop

codons. Promoter and enhancer elements of the SIVcpzUS long terminal repeat (LTR) were indistinguishable from those of other members of the HIV-1/SIVcpz group.

Only three other SIVcpz strains have been reported, two from

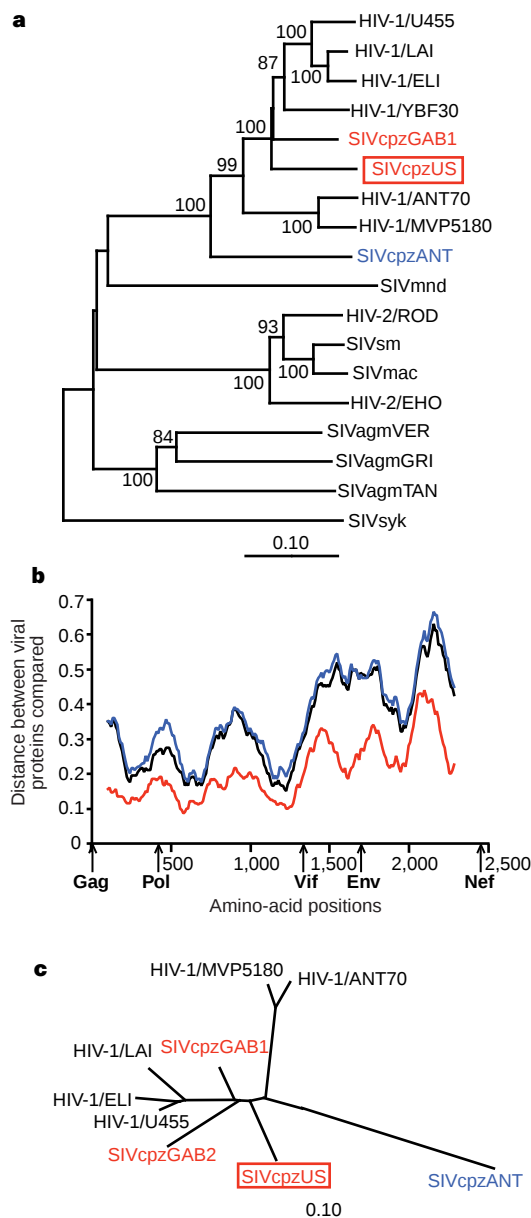


Figure 1 Phylogenetic analysis of SIVcpzUS. **a**, Phylogenetic relationship of SIVcpzUS to other primate lentiviruses. The tree was derived by neighbour-joining analysis²⁷ of full-length *Pol* sequences (trees derived by maximum-likelihood methods²⁸ yielded very similar topologies). Horizontal branch lengths are drawn to scale with the bar indicating 0.1 amino-acid replacements per site. Numbers at each node indicate the percentage of bootstrap samples (out of 1,000) in which the cluster to the right is supported (only values >80% are shown). Other SIVcpz strains closely or more distantly related to SIVcpzUS are shown in red and blue, respectively. **b**, Diversity plots of concatenated SIVcpz protein sequences depicting the proportion of amino-acid sequence differences between SIVcpzUS and SIVcpzGAB1 (red), SIVcpzUS and SIVcpzANT (blue), and SIVcpzGAB1 and SIVcpzANT (black), calculated for a window of 200 amino acids moved in steps of 10 amino acids along the alignment (available as Supplementary Information). The x-axis shows the amino-acid positions along the alignment. The positions of *Gag*, *Pol*, *Vif*, *Env* and *Nef* regions are shown. The y-axis denotes the distance between the viral proteins compared (0.1 = 10% difference). **c**, Unrooted neighbour-joining tree of partial *Pol* protein sequences (distances are drawn to scale).

animals wild-caught in Gabon (SIVcpzGAB1 and SIVcpzGAB2)⁸ and one from a chimpanzee exported to Belgium from Zaire (SIVcpzANT)⁹. SIVcpzGAB1 and SIVcpzANT have been sequenced completely^{1,11}, but only 280 bp of *pol* sequence are available for SIVcpzGAB2 (ref. 10). To determine the evolutionary relationships of SIVcpzUS to these and other HIV and SIV sequences, we performed distance plot and phylogenetic tree analyses using sequences from the HIV sequence database (ref. 14 and [http://](http://hiv-web.lanl.gov/HTML/compendium.html)

hiv-web.lanl.gov/HTML/compendium.html). These analyses identified SIVcpzUS unambiguously as a new member of the HIV-1/SIVcpz group of viruses. A phylogenetic tree of full-length *Pol* sequences showed that SIVcpzUS clustered well within this group but was not particularly closely related to any one human or chimpanzee virus (Fig. 1a). Trees based on other coding regions yielded virtually identical topologies (not shown). Comparison of the phylogenetic position of SIVcpzUS with those of the other SIVcpz strains (Fig. 1a)

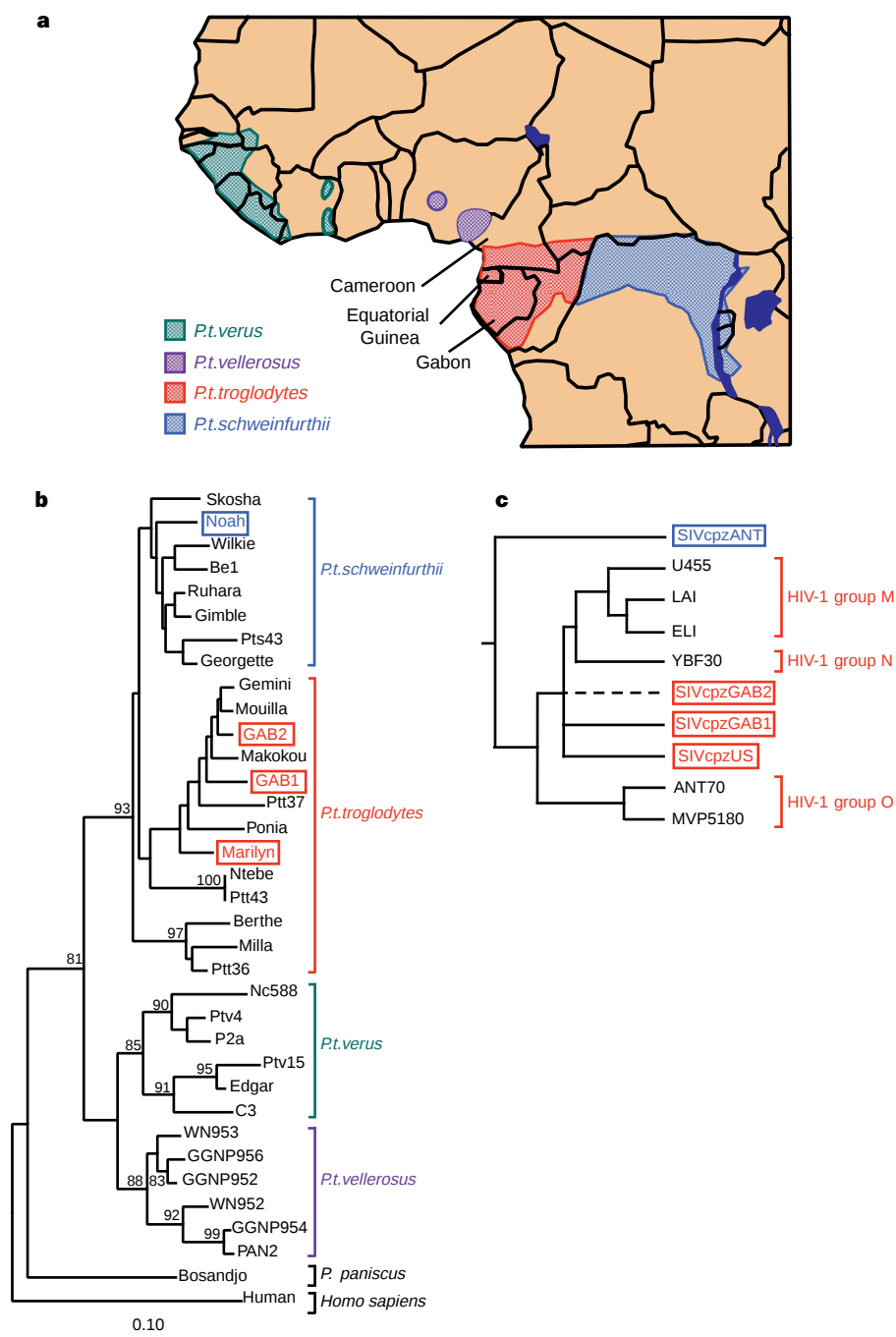


Figure 2 Origin of HIV-1 in *Pan troglodytes troglodytes*. **a**, Geographic ranges of the four subspecies of the common chimpanzee (*Pan troglodytes*) defined by mtDNA analysis (adapted from refs 19, 20 with permission). **b**, Phylogenetic tree of mtDNA sequences. Positions of sequences from the SIVcpz-infected chimpanzees Marilyn (SIVcpzUS), GAB1 (SIVcpzGAB1), GAB2 (SIVcpzGAB2) and Noah (SIVcpzANT) are boxed. The phylogeny was derived by the neighbour-joining method²⁷ applied to pairwise sequence distances calculated using the Kimura two-parameter method (transition/transversion ratio set to 10). Horizontal branch lengths are drawn to scale with the bar indicating 0.1 nucleotide

replacements per site. Numbers at each node indicate the percentage of bootstrap samples (out of 1,000) in which the cluster to the right is supported (only values >80% are shown). Brackets on the right indicate previously defined subspecies/species classifications^{19,20} (*P. t. troglodytes*, *P. t. schweinfurthii*, *P. t. verus*, and *P. t. vellerosus* are colour coded as in **a**). **c**, Schematic tree of *Pol* sequences, highlighting the position of HIV-1 group M, N and O viruses in relation to *P. t. troglodytes* (red) and *P. t. schweinfurthii* (blue) viruses. The position of SIVcpzGAB2 (indicated by broken line), for which only partial sequence is available¹⁰, is inferred from the phylogeny shown in Fig. 1c.

showed that SIVcpzUS was considerably more closely related to SIVcpzGAB1 than to SIVcpzANT. Diversity plots of full-length (concatenated) protein sequences showed that SIVcpzUS was nearly twice as different from SIVcpzANT as from SIVcpzGAB1 (Fig. 1b). When partial Pol sequences of SIVcpzGAB2 were included in phylogenetic analyses (Fig. 1c), SIVcpzANT remained the outlier, differing from the other SIVcpz strains by 23–24%, as compared with only 9–13% divergence among SIVcpzUS, SIVcpzGAB1 and SIVcpzGAB2. These findings indicate that naturally occurring SIVcpz strains fall into two related yet highly divergent phylogenetic lineages.

Divergent lineages of SIV have also been found in African green monkeys^{15–17}. These primates have a broad distribution throughout sub-Saharan Africa and have been classified based on phenotypic differences into four major species, generally known as vervet (*Chlorocebus pygerythrus*), grivet (*C. aethiops*), sabaues (*C. sabaues*) and tantalus (*C. tantalus*) monkeys (note that the genus designation of the African green monkey as *Cercopithecus* has recently been changed to *Chlorocebus*¹⁸). The many SIVagm strains infecting these animals cluster in four distinct phylogenetic lineages according to their species of origin, indicating ancient infection followed by co-evolution of virus and host^{3,15–17}.

To explore whether a similar host-dependent evolution of SIVcpz could account for the extraordinary diversity between SIVcpzANT and the other three SIVcpz strains, we determined the subspecies identity of the animals from which these viruses were derived. Four chimpanzee subspecies with non-overlapping geographic ranges have been proposed on the basis of genetic differences in mitochondrial (mt) DNA sequences^{19,20}. These are the western *P. t. verus*, the Nigerian *P. t. vellerosus*, the central *P. t. troglodytes*, and the eastern *P. t. schweinfurthii* (Fig. 2a). We amplified and sequenced a 498-bp fragment of mitochondrial control region (D-loop) sequences from peripheral-blood mononuclear cell (PBMC) or spleen DNA of the four SIVcpz-infected chimpanzees. Comparison of these newly derived mtDNA sequences to representative sequences from the four chimpanzee subspecies revealed that the three chimpanzees infected with the more closely related SIVcpzGAB1 (GAB1), SIVcpzGAB2 (GAB2), and SIVcpzUS (Marilyn) strains all belonged to the *P. t. troglodytes* subspecies (Fig. 2b). In contrast, the animal infected with the highly divergent SIVcpzANT strain (Noah) was identified as a member of the *P. t. schweinfurthii* subspecies. Classification of the SIVcpz-infected chimpanzees was unambiguous as their mtDNA sequences fell within well-defined subspecies clusters¹⁹ and was further corroborated by the known geographic origins of three of the animals (GAB1, GAB2 and Noah)^{8,9}. We conclude from these results that, as for SIVagm, there has been host-dependent evolution of SIVcpz in chimpanzees.

The discovery of subspecies-specific SIVcpz diversity prompted us to re-examine the phylogenetic positions of all known strains of HIV-1 and SIVcpz, and to look for evidence of cross-species transmission. Globally circulating strains of HIV-1 have been classified into three major phylogenetic groups, termed M, N and O, all of which cause AIDS. The 'main' group M is responsible for the global AIDS epidemic and comprises by far the majority of HIV-1 isolates²¹. These viruses have been further subdivided based on phylogenetic relatedness into ten distinct subtypes or clades, termed A–J (ref. 14). Group O (for 'outlier') is represented by many fewer isolates that originate mainly from Cameroon, Gabon and Equatorial Guinea²². Group N (for 'non-M/non-O') was discovered only very recently²³ and is least widespread of all HIV-1 lineages; so far, it has been documented in only two individuals from Cameroon²³. By comparing the phylogenetic positions of representatives of each of these lineages with those of the four SIVcpz strains, we found that all three HIV-1 groups (M, N and O) clustered closely only with SIVcpz strains infecting chimpanzees of the *P. t. troglodytes* subspecies (Fig. 2c). HIV-1 groups M and N were roughly equidistantly related to SIVcpzGAB1, SIVcpzGAB2 and SIVcpzUS, whereas HIV-1 group O viruses were only slightly more divergent. All SIVcpz strains from

P. t. troglodytes and all three groups of HIV-1 formed a single, monophyletic lineage which was supported by highly significant bootstrap values (>90%). This applied for all coding regions and using different phylogenetic methods (Fig. 1a and data not shown). These data indicate strongly that HIV-1 infection of humans occurred as a result of cross-species transmission of SIVcpz from *P. t. troglodytes*.

Two additional lines of evidence supported a *P. t. troglodytes* origin of HIV-1. First, we found that YBF30, the only fully sequenced example of HIV-1 group N (ref. 23), is a recombinant

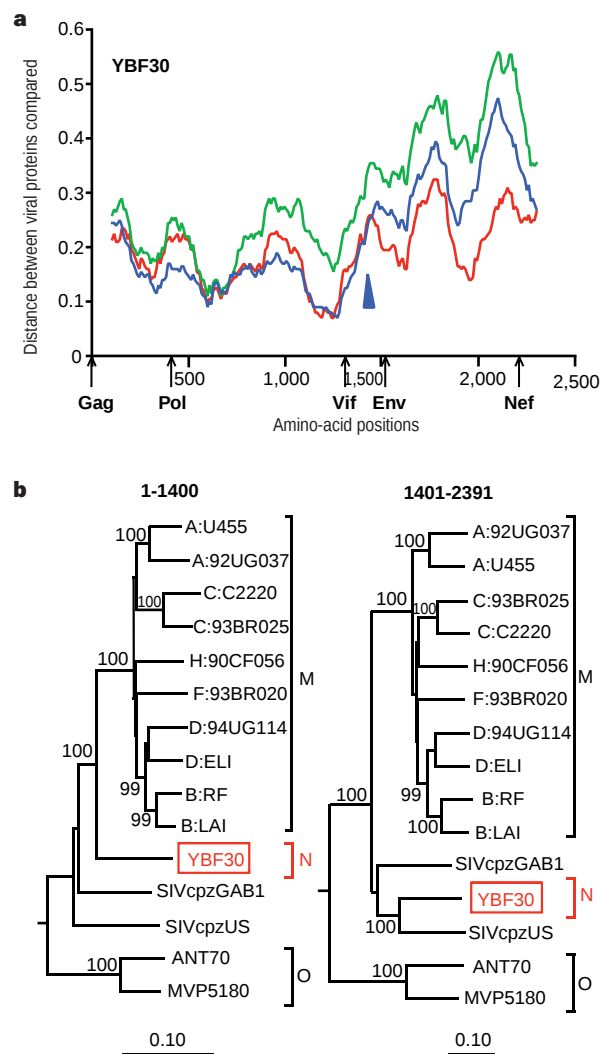


Figure 3 Recombinant origin of HIV-1/YBF30 (group N). **a**, Diversity plots of concatenated protein sequences, depicting the proportion of amino-acid sequence differences between YBF30 (HIV-1 group N; red), U455 (HIV-1 group M; blue), and MVP5180 (HIV-1 group O; green), were calculated for a window of 200 amino acids moved in steps of 10 amino acids along an alignment. The x-axis indicates the amino-acid positions along the alignment. The positions of Gag, Pol, Vif, Env and Nef regions are shown. The y-axis denotes the distance between the viral proteins compared (0.1 = 10% difference). A blue marker at position 1,400 delineates 5' and 3' regions of disproportionate sequence similarity between YBF30 and SIVcpzUS. **b**, Phylogenetic position of YBF30 (boxed) in different parts of its genome. Trees were derived by neighbour-joining analysis²⁷ of concatenated protein sequences flanking the putative recombination breakpoint indicated by the marker in **a** (discordant phylogenies for YBF30 were confirmed by maximum-likelihood methods²⁸). Horizontal branch lengths are drawn to scale with the bar indicating 0.1 amino-acid replacements per site; numbers at each node indicate the percentage of bootstrap samples (out of 1,000) in which the cluster to the right is supported (only values >80% are shown). Brackets identify members of HIV-1 groups M, N and O.

of divergent viral lineages within the HIV-1/SIVcpz(*P.t.t.*) group. Distance plots of full-length (concatenated) protein sequences revealed that YBF30 and SIVcpzUS were disproportionately more similar to each other in the 3' half compared to the 5' half of their genome (Fig. 3a). Phylogenetic tree analyses confirmed these discordant relationships, showing that YBF30 fell into significantly different phylogenetic positions in different parts of its genome (Fig. 3b). For example, in *gag*, *pol* and the 5' half of *vif*, YBF30 sequences formed an independent lineage more closely related to HIV-1 group M than to any SIVcpz; however, in the 3' half of *vif*, and in *env* and *nef*, YBF30 clustered most closely with SIVcpzUS. This mosaic genome structure of YBF30 implies previous co-infection and recombination of divergent SIVcpz strains in a *P. t. troglodytes* host. Second, by carefully analysing three full-length SIVcpz genomes for chimpanzee-specific 'signature' sequences, we found a single protein domain, the V3 loop region of the extracellular envelope glycoprotein, to be conserved uniquely among all SIVcpz strains, even the otherwise highly divergent SIVcpzANT. This sequence conservation was most evident in the V3 crown region which was identical among the three chimpanzee viruses (GPGMTFYN) and differed by only a single amino-acid residue in YBF30 (GPAMTFYN). These data indicate that SIVcpz viruses, all of which share this envelope feature, might as a consequence be uniquely adapted for replication in the chimpanzee host and that YBF30, by virtue of its similarity to SIVcpz in V3, may represent a virus lineage most recently transmitted to humans.

As yet, the oldest trace of the AIDS pandemic is from a human blood sample collected in 1959 from west-central Africa²⁴, although the precise timing and circumstances of early events in the SIVcpz/HIV-1 zoonosis remain obscure. Previous studies exploring the possible origins of HIV-1 made the important discovery that SIVcpz and HIV-1 are isogenic but provided no convincing evidence for chimpanzees—as opposed to another species—as the natural host and reservoir for the human virus¹. In fact, the extent of sequence differences that were observed between HIV-1 and SIVcpz, especially in *vpu*, led to the conclusion that SIVcpz-infected chimpanzees were unlikely to be the proximal source of HIV-1 (ref. 1).

We have shown here that all HIV-1 strains are phylogenetically closely related to SIVcpz strains infecting *P. t. troglodytes*, a primate whose natural range coincides precisely with areas of HIV-1 group M, N, and O endemicity^{13,21–23}. By demonstrating subspecies-dependent evolution of SIVcpz, we also provide evidence for chimpanzees as a long-standing natural reservoir possibly dating to a period before the divergence of *P. t. troglodytes* and *P. t. schweinfurthii*, which is believed to have occurred several hundred thousand years ago¹⁹. The detection of recombination among divergent SIVcpz lineages provides further evidence that SIVcpz infection rates in wild-living chimpanzees must have been (and still may be) substantial; a trivial explanation for the observed low frequency of SIVcpz infection in captive chimpanzees^{8,9,12} is that such animals were either born in captivity or captured as infants before they matured and had increased risk for SIVcpz infection.

Chimpanzees are commonly hunted for food, especially in west equatorial Africa¹³, and as a consequence represent a ready source for zoonotic transmission of SIVcpz to man. Indeed, the phylogenetic positions of HIV-1 groups M, N and O within the HIV-1/SIVcpz radiation (specifically, the interspersed SIVcpz sequences among each of the three HIV-1 groups shown in Fig. 3b, right panel) indicate that the three HIV-1 groups have each arisen as a consequence of independent zoonotic transmissions of SIVcpz from *P. t. troglodytes* to man. Other explanations for the origin of HIV-1 groups M, N and O, such as viral diversification within human populations or acquisition of virus from still another primate species, are either inconsistent with the phylogenetic data or implausible.

We conclude from our results that *P. t. troglodytes* is the natural host and reservoir for HIV-1. It is still possible, however, that the

other chimpanzee subspecies are also infected with SIVcpz (*P. t. schweinfurthii* is an example) and have transmitted their viruses to humans. Such transmissions have not been detected but could have gone unrecognized because of the explosive spread of HIV-1 group M and the absence of serological tests to distinguish SIVcpz(*P.t.t.*) from other SIVcpz lineages. To understand the full extent of natural SIVcpz infection, and the frequency of zoonotic transmission to humans, it will be necessary to screen free-living adult chimpanzees of all four subspecies as well as human populations from corresponding geographic locales. Such studies will require the cooperation of chimpanzee conservationists, local inhabitants and virologists with a keen sensitivity to protection of this endangered species. Notwithstanding these challenges, studies of the natural history and biology of SIVcpz/HIV-1 in chimpanzees and humans promise to yield new insights into the particular circumstances of cross-species transmission and the basis for HIV-1 pathogenicity in humans. □

Methods

PCR amplification and sequence analysis of SIVcpzUS. A complete genomic sequence of SIVcpzUS was obtained by amplifying four overlapping subgenomic fragments from uncultured spleen DNA using nested PCR (see Supplementary Information for details of primer sequences, PCR strategies and amplification conditions). Amplified fragments were cloned and sequenced using the primer walking approach, cycle sequencing and dye terminator methodologies (GenBank accession number AF103818).

PCR amplification and sequence analysis of chimpanzee mitochondrial DNA. A 498-bp segment of the mitochondrial D-loop region (corresponding to positions 15,998–16,497 of the human mitochondrial sequence²⁵) was amplified from PBMC (GAB1, GAB2, Noah) or spleen DNA (Marilyn) using single-round PCR (see Supplementary Information) and sequenced without interim cloning (GenBank accession numbers: GAB1/AF102683; GAB2/AF102684; Marilyn/AF102685; Noah/AF102687).

Sequence comparisons. SIVcpzUS-predicted protein sequences were aligned with those of other HIV and SIV reference strains¹⁴ using CLUSTAL_X (ref. 26). Complete proteome alignments were constructed by concatenating *Gag*, *Pol*, *Vif*, *Env* and *Nef* alignments (available as Supplementary Information); in the regions of *gag-pol* and *pol-vif* gene overlap, the *Gag* and *Pol* sequences, respectively, were excluded. Phylogenetic analyses were done using neighbour-joining and maximum-likelihood methods. The neighbour-joining method²⁷ was applied to protein-sequence distances calculated by the method of Kimura, with 1,000 bootstrap replicates, as implemented in CLUSTAL_X. The maximum-likelihood method used the JTT model of amino-acid replacement, was replicated five times with shuffled input order, and was implemented in MOLPHY²⁸. MtDNA reference sequences were derived from refs 19, 20.

Received 9 November; accepted 21 December 1998.

- Huet, T., Cheynier, R., Meyerhans, A., Roelants, G. & Wain-Hobson, S. Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* **345**, 356–359 (1990).
- Hirsch, V. M., Olmsted, R. A., Murphy-Corb, M., Purcell, R. H. & Johnson, P. R. An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature* **339**, 389–392 (1989).
- Sharp, P. M., Robertson, D. L. & Hahn, B. H. Cross-species transmission and recombination of AIDS viruses. *Phil. Trans. R. Soc. Lond. B* **349**, 41–47 (1995).
- Gao, F. et al. Human infection by genetically-diverse SIVsm-related HIV-2 in west Africa. *Nature* **358**, 495–499 (1992).
- Gao, F. et al. Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology. *J. Virol.* **68**, 7433–7447 (1994).
- Chen, Z. et al. Genetic characterization of a new west African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *J. Virol.* **70**, 3617–3627 (1996).
- Chen, Z. et al. Human immunodeficiency virus type 2 (HIV-2) seroprevalence and characterization of a distinct HIV-2 genetic subtype from the natural range of SIV infected sooty mangabeys. *J. Virol.* **71**, 3953–3960 (1997).
- Peeters, M. et al. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS* **3**, 625–630 (1989).
- Peeters, M. et al. Isolation and characterization of a new chimpanzee lentivirus (simian immunodeficiency virus isolate cpz-ant) from a wild-captured chimpanzee. *AIDS* **6**, 447–451 (1992).
- Janssens, W. et al. Phylogenetic analysis of a new chimpanzee lentivirus SIVcpz-gab2 from a wild-captured chimpanzee from Gabon. *AIDS Res. Hum. Retroviruses* **10**, 1191–1192 (1994).
- Vanden Haesevelde, M. M. et al. Sequence analysis of a highly divergent HIV-1 related lentivirus isolated from a wild captured chimpanzee. *Virology* **221**, 346–350 (1996).
- Gilden, R. V. et al. HTLV-III antibody in a breeding chimpanzee not experimentally exposed to the virus (letter). *Lancet* **i**, 678–679 (1986).

Correspondence and requests for materials should be addressed to B.H.H. (e-mail: bhahn@uab.edu).

Molecular characterization of mitochondrial apoptosis-inducing factor

**Santos A. Susin^{*}, Hans K. Lorenzo[†], Naoufal Zamzami^{*},
Isabel Marzo^{*}, Bryan E. Snow[‡], Greg M. Brothers[‡],
Joan Mangion[‡], Etienne Jacotot^{*}, Paola Costantini^{*},
Markus Loeffler^{*}, Nathanael Larochette^{*},
David R. Goodlett[§], Ruedi Aebersold[§],
David P. Siderovski[‡], Josef M. Penninger[‡] & Guido Kroemer^{*}**

* Centre National de la Recherche Scientifique, UPR 420, 19 rue Guy Môquet, F-94801 Villejuif, France

† Unité de Biochimie Structurale, Institut Pasteur, 25 rue du Dr Roux,
F-75724 Paris Cedex 15, France

‡ The Amgen Institute and Ontario Cancer Institute, Department of Medical Biophysics and Immunology, University of Toronto, 620 University Avenue, Suite 706, Toronto, Ontario M5G 2C1, Canada

§ Department of Molecular Biotechnology, University of Washington, Seattle, Washington 98195, USA

Mitochondria play a key part in the regulation of apoptosis (cell death)^{1,2}. Their intermembrane space contains several proteins that are liberated through the outer membrane in order to participate in the degradation phase of apoptosis³⁻⁹. Here we report the identification and cloning of an apoptosis-inducing factor, AIF⁵, which is sufficient to induce apoptosis of isolated nuclei. AIF is a flavoprotein of relative molecular mass 57,000 which shares homology with the bacterial oxidoreductases; it is normally confined to mitochondria but translocates to the nucleus when apoptosis is induced. Recombinant AIF causes chromatin condensation in isolated nuclei and large-scale frag-

mentation of DNA. It induces purified mitochondria to release the apoptogenic proteins cytochrome *c* and caspase-9. Microinjection of AIF into the cytoplasm of intact cells induces condensation of chromatin, dissipation of the mitochondrial transmembrane potential, and exposure of phosphatidylserine in the plasma membrane. None of these effects is prevented by the wide-ranging caspase inhibitor known as Z-VAD.fmk. Overexpression of Bcl-2, which controls the opening of mitochondrial permeability transition pores, prevents the release of AIF from the mitochondrion but does not affect its apoptogenic activity. These results indicate that AIF is a mitochondrial effector of apoptotic cell death.

Opening of the mitochondrial permeability transition pore, which is under the control of members of the Bcl-2 family, is one of the decisive events of the apoptotic process^{1,2}, causing an increase in the permeability of the outer mitochondrial membrane⁷ and the release of soluble proteins from the intermembrane space. The mitochondrial intermembrane fraction contains several potentially apoptogenic factors, including cytochrome *c* (refs 4, 6), procaspases 2, 3 and 9 (refs 8, 9), and AIF, which forces isolated nuclei to adopt an apoptotic morphology^{3,5}. We purified an AIF activity that is maintained in the presence of the caspase inhibitor

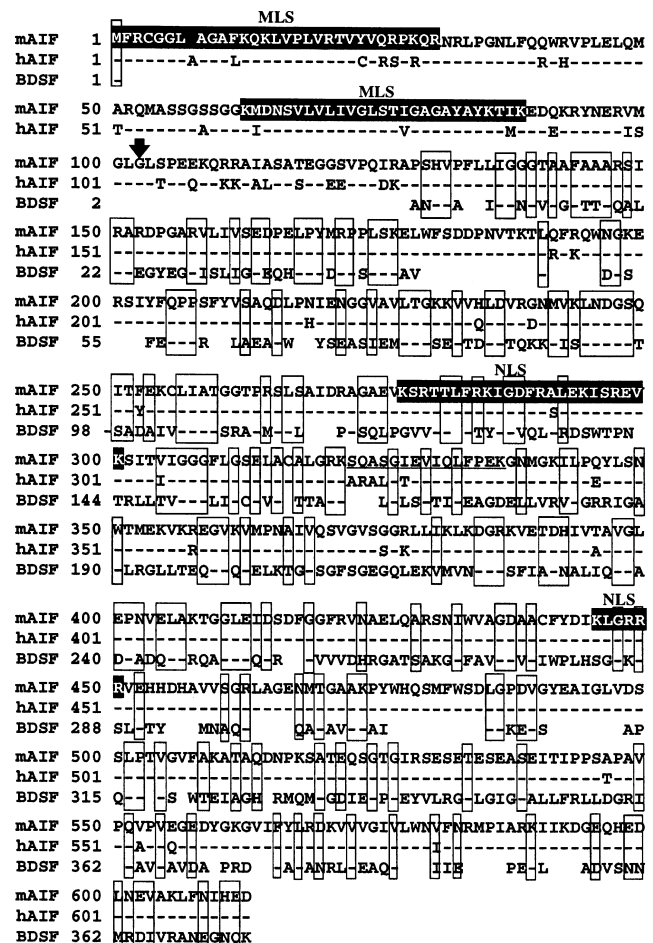


Figure 1 Alignment of mouse and human AIF amino-acid sequences with benzene 1,2-dioxygenase system ferredoxin NADH reductase from *Pseudomonas putida* (BDSF; 30% amino-acid identity with mouse AIF). Dashes indicate amino-acid identity; lined boxes indicate amino-acid similarity; black boxes highlight mitochondrial localization sequences (MLS; ref. 11) and putative nuclear-localization sequences (NLS; ref. 13). Arrow, first residue (amino acid 102) of mature AIF purified from mouse liver, as determined by N-terminal sequencing. The underlined sequence in mouse AIF matches the mass spectroscopy data obtained with trypsin-digested purified AIF. GenBank accession numbers for mouse and human AIF are AF100927 and AF100928, respectively.

Z-VAD.fmk (ref. 9) from the supernatant of mouse liver mitochondria that had their permeability transition pores open. Tandem mass spectrometry of a trypsin digest of material extracted from a single silver-stained SDS-polyacrylamide gel electrophoresis (PAGE) band¹⁰ was used to identify an expressed sequence tag (EST; GenBank no. 1595214) whose amino-acid sequence is underlined in Fig. 1; this enabled us to clone the corresponding full-length complementary DNAs from mouse and human (Fig. 1). AIF is strongly conserved between the two mammalian species (92% amino-acid identity) and its carboxy-terminal portion (residues 128 to 612 for mouse AIF) shares significant homology with several bacterial ferredoxin or NADH-oxidoreductases (but there is no homology with known components of the respiratory chain). AIF's amino-terminal portion bears a mitochondrial localization sequence¹¹ (boxed in Fig. 1). Only one mouse chromosome hybridizes with AIF cDNA *in situ* (results not shown). This gene lies within the X-chromosome region A6 in the mouse, which is syntenic to the human X-chromosome region Xq25–26 where the AIF gene is located (EMBL accession no. Z81364).

An antibody raised against the amino-acid residues 151 to 200 of AIF recognizes a single protein of relative molecular mass ~57K in mitochondria from mouse liver (Fig. 2a) and many different tissues (results not shown). In contrast, the primary transcription/

translation product of AIF cDNA *in vitro* is ~67K. When imported into mitochondria *in vitro*, this gives rise to a shorter protein (Fig. 2b) corresponding to mature AIF (whose first amino acid is at position 102 of the full-length precursor protein; Fig. 1). There was no import if residues 1–120 of the protein were deleted, which removes the mitochondrial presequence (not shown). AIF activity and immunoreactivity are contained inside the mitochondrial intermembrane space (Fig. 2a) until they are released upon opening of the permeability transition pores by atractyloside (Atr), Ca²⁺ or *tert*-butylhydroperoxide. This release is inhibited by the effect of cyclosporin A (CsA) on the pores (Fig. 2a). Immunodepletion of AIF from the pool of intermembrane proteins also removes the activity that induces nuclear apoptosis *in vitro*, indicating that AIF is the principal mitochondrial factor causing nuclear apoptosis (Fig. 2a). Subcellular fractionation (Fig. 2c) and immunofluorescence analysis (Fig. 2d) confirm that AIF is confined to mitochondria in healthy cells. After induction of apoptosis by staurosporine, AIF translocates to the cytosol and to the nucleus (Fig. 2d) at the same time as the mitochondrial transmembrane potential ($\Delta\Psi_m$) is reduced and nuclei become translucent with Hoechst 33342 staining and show peripheral condensation of their chromatin (stage I in Fig. 2d–f). In Rat-1 cells, full release of cytochrome *c* from the mitochondrion to the cytosol is only visible at a subsequent stage,

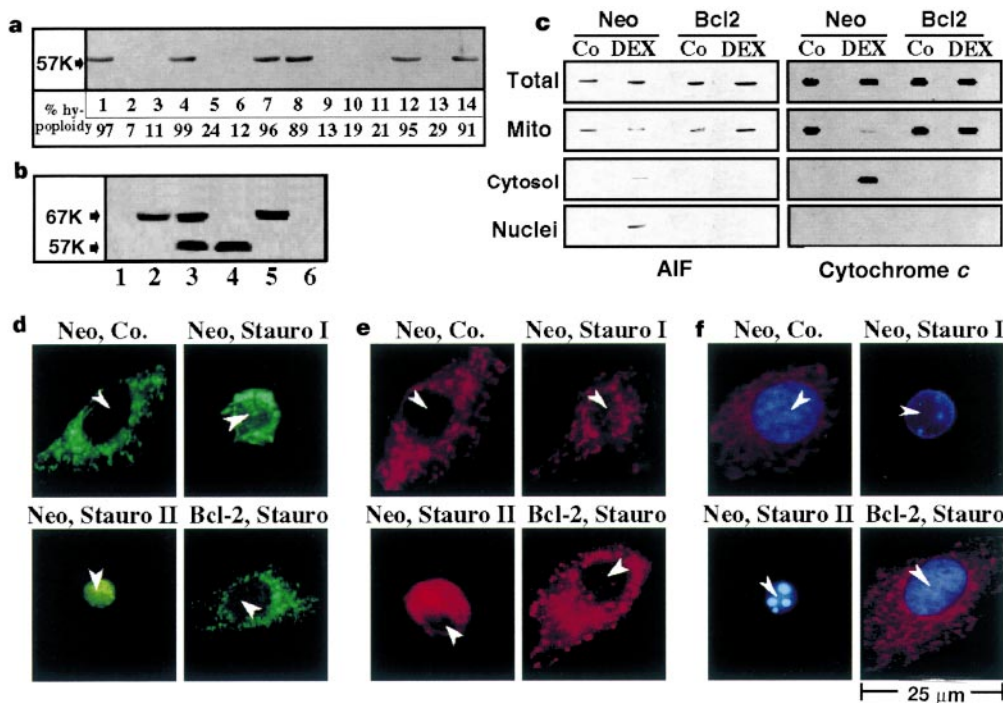


Figure 2 Subcellular and submitochondrial distribution of AIF. **a**, Submitochondrial localization of AIF. Western blots were performed on total mouse liver mitochondrion lysate (lane 1), proteins for the matrix (lane 2), inner membrane (lane 3), intermembrane space (lane 4), outer membrane (lane 5), supernatant from untreated mitochondria (lane 6), or from mitochondria treated with 5 mM Atr (lane 7), 200 μ M Ca²⁺ (lane 8), 1 μ M CsA (lane 9), Atr + CsA (lane 10) or Ca²⁺ + CsA (lane 11). In addition, the supernatant of Atr-treated mitochondria was sham-immunodepleted using a preimmune serum (lane 12), or AIF-immunodepleted in the absence (lane 13) or presence (lane 14) of AIF-derived peptides (5 μ M) blocking the AIF antiserum. Aliquots of each preparation were tested for their capacity to induce nuclear hypoploidy in isolated HeLa nuclei. **b**, Mitochondrial import of AIF. *In vitro* transcription and translation were done using the TNT lysate-coupled transcription/translation kit (Promega) and [³⁵S]methionine in the absence (lane 1) or presence (lane 2) of mouse AIF cDNA. The retention of this product was measured using mitochondria that were left untreated (lane 3), digested with proteinase K to remove surface-bound protein (lane 4), de-energized with 100 μ M CCCP (lane 5) or treated with both proteinase K and CCCP (lane

6)²⁹. Note the maturation of AIF to a 57K protein which is retained by mitochondria in a CCCP-inhibitable fashion and is protected from proteolysis. **c**, Subcellular localization of AIF as compared to cytochrome *c*. 2B4.11 T-cell hybridoma expressing a control vector (Neo) or human Bcl-2 (ref. 5) were cultured with dexamethasone (DEX, 1 μ M; for 12 h), followed by immunoblot detection of AIF or cytochrome *c* in total cell lysates or subcellular fractions⁴. Adequate separation of mitochondria was tested using anti-Hsp60 (not shown). Co, control. **d**, Immunofluorescence detection of AIF in Rat-1 fibroblasts transfected with a vector control (Neo) or Bcl-2 (ref. 30) and treated with staurosporine (stauro; 1 μ M; 2 h) or left untreated (Co). In Neo cells, staurosporine has two effects, either causing partial chromatin condensation (~80% of cells; stage I) or advanced chromatin condensation and nuclear fragmentation (~20% of cells; stage II), as identified by phase contrast microscopy or counterstaining with Hoechst 33342 (results not shown). Arrowheads indicate the centre of the nucleus. **e**, Immunofluorescence detection of cytochrome *c*; cells were treated and classified as in **d**. **f**, Nuclear morphology (Hoechst 33342, blue) and $\Delta\Psi_m$ (CMXRos, red) of live cells.

namely when chromatin condensation and nuclear fragmentation (karyorrhexis) are advanced (stage II in Fig. 2d–f) and AIF concentrates in the nucleus (Fig. 2c). The nuclear localization of AIF is compatible with its overall amino-acid composition¹² and the presence of putative nuclear-localization signals (Fig. 1)¹³. Overexpression of Bcl-2 impedes the staurosporine-triggered mitochondrial release of AIF and cytochrome *c* and stabilizes the $\Delta\Psi_m$ (Fig. 2c–f), in agreement with previous observations^{3–7,9}. The

differential relocalization of the two intermembrane proteins cytochrome *c* and AIF, which translocate to the cytosol and to the nucleus, respectively, is confirmed in other experimental models of apoptosis, including T-cell hybridoma cells treated with dexamethasone (Fig. 2c).

When added to purified nuclei from HeLa cells, recombinant AIF induces loss of DNA (Fig. 3a) and peripheral condensation of chromatin, as determined by 4,6-diamidino-2-phenylindole

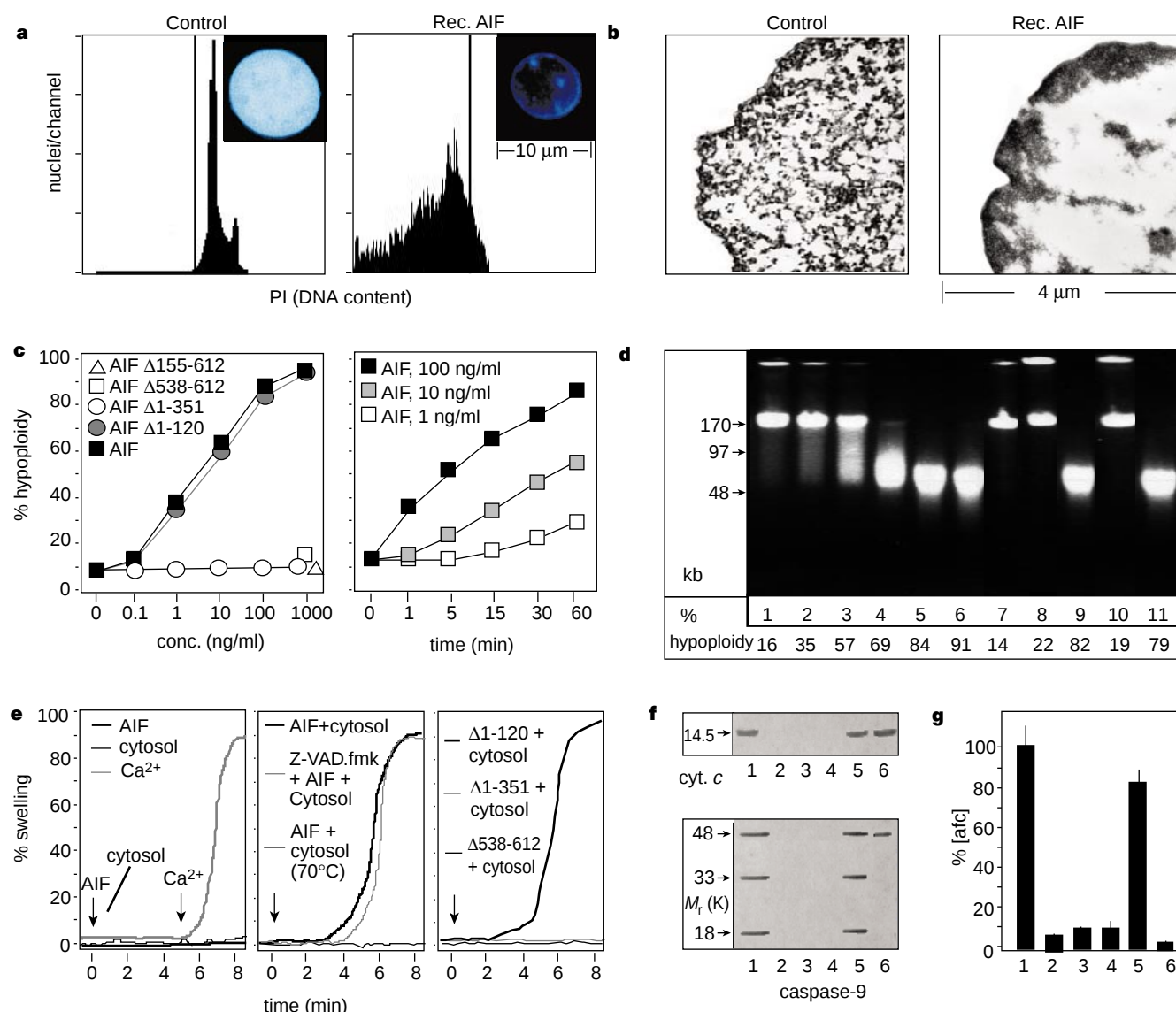


Figure 3 Effects of AIF on isolated nuclei and mitochondria. **a**, AIF-induced DNA loss and chromatin condensation in isolated nuclei. HeLa nuclei were cultured at 37°C in the absence (control) or presence of 100 ng ml⁻¹ recombinant AIF for 90 min, stained with propidium iodide (PI), and analysed by flow cytometry. Inserts show typical results of DAPI staining. **b**, Electron-microscopic determination of chromatin condensation of cells treated as in **a**. **c**, Concentration and time dependence of AIF effects on isolated nuclei, determined as for **a**. Nuclei were cultured for 90 min with recombinant AIF or deletion mutants (left), or for different periods with AIF (right). **d**, Pulse-field gel electrophoresis of HeLa nuclei cultured for 0 min (lane 1), 5 min (lane 2), 15 min (lane 3), 30 min (lane 4), 60 min (lane 5) or 90 min (lane 6) with 100 ng ml⁻¹ AIF alone, for 90 min with 100 ng ml⁻¹ AIFΔ1–351 (lane 7), or for 90 min with 100 ng ml⁻¹ AIF (lanes 8–11) in the presence of 200 μ M *p*-chloromercuriphenylsulphonic acid (lane 8), 200 μ M Z-VAD.fmk (lane 9), 5 mM EDTA (lane 10) or 5 mM EGTA (lane 11). Each nuclear preparation was also assessed for hypodiploidy (per cent) as in **a**. **e**, Swelling of purified rat mitochondria induced by AIF. Arrows indicate points of addition of Ca²⁺ (200 μ M;

positive control), AIF or AIF mutants (100 ng ml⁻¹), cytosol (100 μ g protein⁻¹, optionally inactivated at 70°C for 5 min), and/or Z-VAD.fmk (100 μ M, added beforehand to the cytosol/AIF mixture). Note that only the combined addition of AIF plus cytosol (or AIFΔ1–120 plus cytosol) causes mitochondrial swelling. **f**, AIF-induced release of cytochrome *c* and caspase-9 from mitochondria. Isolated mitochondria were subjected to osmotic lysis (lane 1) or treated as in **a** (15 min) with buffer only (lane 2), recombinant AIF (lane 3), cytosol (lane 4), recombinant AIF + cytosol (lane 5), or recombinant AIF + cytosol + Z-VAD.fmk (lane 6), and their supernatants were subjected to immunoblot analysis of cytochrome *c* and caspase-9. Note that Z-VAD.fmk does not inhibit the release of caspase-9 but rather interferes with its proteolytic activation. **g**, AIF-induced activation of a caspase. The same mitochondrial supernatants as in **f** were tested for their ability to cleave the caspase substrate Z-VAD.afc. The 100% value refers to the enzyme activity obtained by osmotic lysis of mitochondria (lane 1).

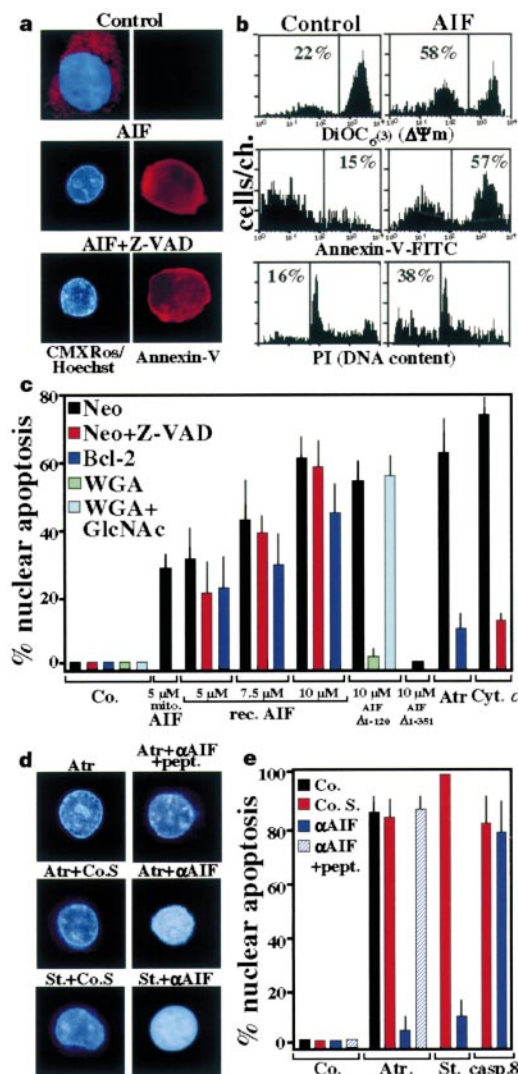


Figure 4 Effects of AIF on intact cells. **a**, Microinjection of AIF. Buffer only (Co, control) or recombinant AIF (10 μ M) was injected into the cytoplasm of Rat-1 cells, which were cultured for 90 min in the absence or presence of 100 μ M Z-VAD.fmk (added 15 min before injection). Representative (three independent experiments, 100–200 microinjected cells per session) microphotographs of viable cells stained with CMXRos and Hoechst 33342 (as in Fig. 3f) (left panels) or annexin V, which labels surface-exposed phosphatidylserine residues (red fluorescence in right panels). **b**, Apoptosis induced by transfection-enforced AIF overexpression. Jurkat cells were transiently transfected with pcDNA3.1 vector only (control) or mouse AIF cDNA. Parallel transfections with a Myc-tagged AIF revealed a transfection efficiency of 40–50%. After 24 hours in culture, apoptosis-associated alterations were measured by cytofluorimetry. **c**, Quantification of nuclear apoptosis induced by microinjection (≥ 100 cells; 2–3 determinations; values are means $\pm$ s.e.m.) of purified⁹ mitochondrial (mito) AIF, recombinant AIF, cytochrome *c*, WGA, WGA + *N*-acetylglucosamine (GlcNAc) and/or Atr into Neo- or Bcl-2-transfected Rat-1 cells or Neo cells treated with Z-VAD.fmk. Nuclear apoptosis is accompanied by $\Delta\Psi_m$ dissipation, including in Bcl-2 overexpressing cells. **d**, Inhibition of Atr- and staurosporin-induced nuclear apoptosis by anti-AIF antiserum. Rat-1 cells were microinjected with Atr (50 μ M) alone, or Atr diluted in control serum, anti-AIF (α AIF) antiserum, and/or 100 μ M AIF-derived immunogenic peptides. 180 min after microinjection, cells were stained with Hoechst 33342 and CMXRos. Alternatively, cells were cultured for 120 min with staurosporin (1 μ M, as in Fig. 2d–f), added to the culture medium after microinjection of control antiserum or anti-AIF. Atr-injected and staurosporin-treated cells do not retain the $\Delta\Psi_m$ -sensitive dye CMXRos. **e**, Quantification of nuclear apoptosis inhibited by anti-AIF antibody. Cells were microinjected with Atr, control serum, anti-AIF, and/or AIF-derived peptides, and/or treated with staurosporin (St.), as in **d**. Alternatively, cells were microinjected with recombinant active caspase-8 (1 U μ L⁻¹ for 180 min).

(DAPI) staining (Fig. 3a, inserts) and confirmed by electron microscopy (Fig. 3b). These effects are seen after addition of not only the entire recombinant AIF-precursor protein, but also for a deletion mutant lacking the mitochondrial presequence (Δ 1–120); they are not evident after addition of deletion mutants Δ 1–351, Δ 155–612 or Δ 538–612. Only the deletion mutant Δ 1–120 spontaneously binds FAD (needed for the oxidoreductase function), thereby behaving like native mitochondrial AIF, a flavoprotein. The recombinant AIF-precursor protein, however, does not contain FAD when purified from bacteria. Recombinant AIF-precursor protein becomes apoptogenic after refolding on a nickel-affinity matrix even in the absence of FAD, indicating that any putative AIF oxidoreductase activity is unnecessary for its apoptogenic function. This function of AIF seems to be conformation-dependent in that refolding is required, which is readily destroyed by trichloric acid, mild heat (65 °C), or precipitation with ammonium sulphate (not shown). The effect of AIF on isolated nuclei does not depend on other cytoplasmic factors and is rapid (within 1 min; Fig. 3c), with concomitant digestion of chromatin into fragments of $\sim$ 50 kilobases (Fig. 3d). This large-scale DNA fragmentation is inhibited by the Mg²⁺ chelator EDTA and by the thiol-reactive *p*-chloromercuriphenylsulphonic acid, but not by the caspase inhibitor Z-VAD.fmk (Fig. 3d). Recombinant AIF added to purified nuclei does not cause oligonucleosomal fragmentation of DNA, nor does it cleave purified plasmid DNA (results not shown). In the presence of a heat-labile cytosolic factor, AIF (and deleted versions of the protein: AIF Δ 1–120 but not Δ 1–351, Δ 155–612 or Δ 538–612) causes purified mitochondria to swell markedly, which is indicative of an increase in the permeability of the mitochondrial membrane (Fig. 3e); this effect is accompanied by the escape of cytochrome *c* and caspase-9 (Fig. 3f). None of these AIF effects, either on isolated nuclei or on mitochondria, is prevented by Z-VAD.fmk (Fig. 3d–f), suggesting that they do not depend on caspase. However, the supernatant of mitochondria treated with AIF-plus-cytosol contains a Z-VAD.fmk-inhibitable enzymatic activity which cleaves the caspase substrate Z-VAD.afc (Fig. 3g). This activity is at least partly due to the presence of activated caspase-9 (Fig. 3f, and ref. 9). Thus, AIF may activate caspase-9 (and presumably other caspases) through an indirect, mitochondrion-dependent mechanism.

When microinjected into the cytoplasm of live cells, recombinant AIF and AIF Δ 1–120 (but not Δ 1–351, Δ 155–612 or Δ 538–612) rapidly (60–90 min) induce several events that are hallmarks of apoptosis: nuclear chromatin condensation and DNA loss, dissipation of the $\Delta\Psi_m$, and exposure of phosphatidylserine on the outer leaflet of the plasma membrane (Fig. 4a, c). The amount of AIF required for this effect (5–10 μ M; Fig. 4c; approximate injection volumes of 20–200 femtolitres per cell) is within the physiological range of the total AIF content per cell (1–3 attomol per cell in Rat-1 cells), as estimated by semiquantitative immunoblots. Transfection-enforced overexpression of wild-type AIF also induces the collapse of $\Delta\Psi_m$, phosphatidylserine exposure, and hypoploidy (Fig. 4b). None of the effects mediated by microinjected AIF is inhibited by Z-VAD.fmk (Fig. 4a, c), although Z-VAD.fmk succeeds in preventing apoptosis that is induced by cytochrome *c* and is dependent on caspase^{6,14} (Fig. 4c). Careful titration of AIF revealed no significant difference in its efficacy to induce nuclear apoptosis (Fig. 4c) and $\Delta\Psi_m$ dissipation (not shown) in control cells and in Bcl-2-transfected cells. As an internal control for its cytoprotective effect, Bcl-2 prevents the loss of $\Delta\Psi_m$ (refs 3, 5) and the nuclear apoptosis induced by microinjection of the permeability-transition-pore-opening agent Atr (Fig. 4c). Inhibition of active nuclear transport by co-injection of the lectin wheat-germ agglutinin¹⁵ does prevent AIF-induced nuclear apoptosis (Fig. 4c) without affecting AIF-induced $\Delta\Psi_m$ dissipation results (not shown). As a control, blockade of the lectin moiety by *N*-acetylglucosamine abolishes the wheat-germ-agglutinin-mediated inhibition of AIF-induced nuclear apoptosis (Fig. 4c). These results indicate that AIF operates

independently of Bcl-2 and caspases, yet requires energy-dependent transport into the nucleus to exert local effects.

Microinjection of the AIF-specific antiserum abolishes morphological signs of Atr-induced nuclear apoptosis (Fig. 4d, e), although it does not impede Atr-induced $\Delta\Psi_m$ dissipation (Fig. 4d) and later cell death. There is no inhibition when AIF-specific antibody is neutralized by co-injection of an excess of immunogenic peptides. Anti-AIF antiserum also prevents nuclear changes induced by staurosporine (Fig. 4d, e), underscoring the contribution of AIF to at least some pathways culminating in nuclear apoptosis. In contrast, anti-AIF antiserum has no effect on nuclear apoptosis induced by microinjection of recombinant caspase-8 (Fig. 4e), indicating the existence of a caspase-dependent, AIF-independent (and presumably mitochondrion-independent)^{16,17} pathway that leads to nuclear apoptosis.

Our results show that AIF is an apoptogenic mitochondrial intermembrane protein. Like cytochrome *c*, AIF is likely to be a phylogenetically old, bifunctional protein with an electron acceptor/donor (oxidoreductase) function and a second, independent apoptogenic function¹⁸. Cytochrome *c* is assembled in the intermembrane space from two non-apoptogenic precursor molecules, haem and apocytochrome *c*, in a reaction that allows the molecule to adopt an apoptogenic conformation¹⁸. Likewise, the primary *in vitro* transcription/translation product of AIF is not apoptogenic (not shown), and the acquisition of AIF activity depends on folding, which may be favoured by mitochondrial import, as for other flavoproteins¹⁹. Cytochrome *c* redistributes from the mitochondrion to the cytosol and induces nuclear apoptosis with the help of additional molecules (Apaf-1, ATP and procaspase-9), which together activate caspase-3, thereby activating another factor, DFF/CAD, which triggers oligonucleosomal DNA fragmentation^{14,16,20}. Although DFF/CAD is an important cytoplasmic effector of nuclear apoptosis^{16,20}, it is not the only one responsible for apoptotic chromatin condensation²¹. In contrast to cytochrome *c*, AIF has a direct effect on isolated nuclei, triggering chromatin condensation as well as large-scale chromatin fragmentation. This type of DNA fragmentation precedes oligonucleosomal DNA degradation in several cellular models of apoptosis^{22,23} and can be caspase-independent²⁴ (Fig. 3d). AIF also affects the barrier function of mitochondrial membranes (Fig. 3e, f), suggesting that it can engage in a self-amplification loop whereby AIF released from some mitochondria acts on other mitochondria to compromise their membrane function²⁵. Bcl-2 inhibits the mitochondrial release of AIF (Fig. 2b, c) but has no cytoprotective effect once AIF is present in the cytosol (Fig. 4c). Thus it is likely that AIF acts beyond or independently of the Bcl-2 and caspase checkpoints in the cell-death process. We conclude that AIF provides a new molecular link between mitochondrial membrane permeabilization and nuclear apoptosis, and that the caspases, DFF/CAD and AIF are probably engaged in complementary cooperative or redundant pathways that lead to nuclear apoptosis. □

Methods

Recombinant AIF protein. Thioredoxin-tagged mouse AIF (amino acids 1–612; numbering as for Fig. 1) and AIF deletion mutants ($\Delta 1$ –120, $\Delta 1$ –351, $\Delta 155$ –612, $\Delta 538$ –612) were expressed from a Novagen pET32 expression vector and purified from *Escherichia coli*. Proteins were refolded on a nickel–NTA affinity matrix and stored in 50 mM HEPES, pH 7.9, 100 mM NaCl, 2 mM EDTA, 1 mM DTT and 10% glycerol.

AIF antiserum, immunodetection and immunodepletion. A rabbit antiserum was generated against a mixture of three AIF peptides (amino acids 151–170, 166–185, 181–200, coupled to keyhole limpet haemocyanin) and was used in immunoblots (1/2,000) and on fixed²⁶ Rat-1 cells (1/250) with a goat anti-rabbit IgG conjugated to peroxidase or FITC, respectively. Cytochrome *c* was detected by immunofluorescence²⁶ in fixed Rat-1 cells by a monoclonal antibody revealed with a secondary phycoerythrin-labelled antibody. The $\Delta\Psi_m$ -sensitive dye CMXRos (100 nM) was used on live cells²⁷

counterstained with the Hoechst 33342 dye (1 μ M). Immunodepletion of AIF from intermembrane proteins (100 μ g ml^{–1}) was achieved by anti-AIF immobilized on protein A/G–agarose beads (Santa Cruz Biotechnology).

Cell-free systems of apoptosis. Purified HeLa cell nuclei were exposed to AIF in CFS buffer²⁵, and nuclear apoptosis was quantified by staining with the DNA-intercalating dye propidium iodide and by cytofluorometric determination of DNA content²⁵. Isolated rat liver mitochondria (0.5 mg protein ml^{–1}) were exposed to cytosol (100 μ g protein ml^{–1})^{25,28} and/or recombinant AIF (100 ng ml^{–1}), while monitoring for large-amplitude swelling at OD540. The release of cytochrome *c* and caspase-9 into the supernatant was measured by immunoblot using a monoclonal anti-cytochrome *c* antibody (7H8.2C12, Pharmingen), or a rabbit antibody against the large subunit of caspase-9 (Hazelton).

Microinjection, transfection and quantification of apoptosis. Rat-1 fibroblast cells were microinjected with buffer only, Atr (50 μ M) dialysed antiserum, AIF-derived peptides (100 μ M), horse cytochrome *c* (30 μ M; Sigma), WGA (1 mg ml^{–1}; Sigma), *N*-acetylglucosamine (3 mM), and/or recombinant AIF. Z-VAD.fmk (100 μ M; Bachem) was added to the culture medium 30 min before microinjection into the cytoplasm (pressure, 150 hPa; 0.2 s; ref. 27). After microinjection, cells were cultured for 90–180 min and stained for 10 min with CMXRos, Hoechst 33342 dye, or annexin V conjugated to biotin (Boehringer) and revealed by an avidin–phycoerythrin conjugate (Sigma). Microinjected viable cells (100–200 per session, two to three independent sessions of injection) were identified by inclusion of 0.25% (w/v) FITC–dextran (green fluorescence) in the injectate (not shown). Only the blue or red fluorescence was recorded. Jurkat T-lymphoma cells were transfected using AIF cloned in pcDNA3.1 (Invitrogen) vector and Lipofectamine (Gibco Life Technologies).

Received 4 November; accepted 21 December 1998.

- Kroemer, G. The proto-oncogene Bcl-2 and its role in regulating apoptosis. *Nature Med.* **3**, 614–620 (1997).
- Green, D. R. & Reed, J. C. Mitochondria and apoptosis. *Science* **281**, 1309–1312 (1998).
- Zamzami, N. *et al.* Mitochondrial control of nuclear apoptosis. *J. Exp. Med.* **183**, 1533–1544 (1996).
- Liu, X. S., Kim, C. N., Yang, J., Jemmerson, R. & Wang, X. Induction of apoptotic program in cell-free extracts: requirement for dATP and cytochrome *c*. *Cell* **86**, 147–157 (1996).
- Susin, S. A. *et al.* Bcl-2 inhibits the mitochondrial release of an apoptogenic protease. *J. Exp. Med.* **184**, 1331–1342 (1996).
- Kluck, R. M., Bossy-Wetzel, E., Green, D. R. & Newmeyer, D. D. The release of cytochrome *c* from mitochondria: a primary site for Bcl-2 regulation of apoptosis. *Science* **275**, 1132–1136 (1997).
- Vander Heiden, M. G., Chandal, N. S., Williamson, E. K., Schumacker, P. T. & Thompson, C. B. Bcl-XL regulates the membrane potential and volume homeostasis of mitochondria. *Cell* **91**, 627–637 (1997).
- Mancini, M. *et al.* The caspase-3 precursor has a cytosolic and mitochondrial distribution: Implications for apoptotic signaling. *J. Cell Biol.* **140**, 1485–1495 (1998).
- Susin, S. A. *et al.* Mitochondrial release of caspases-2 and -9 during the apoptotic process. *J. Exp. Med.* **189**, 381–394 (1999).
- Ducret, A., Van Oostveen, L., Eng, J. K., Yates, J. R. & Aebersold, R. High-throughput protein characterization by automated reverse-phase chromatography electrospray tandem mass spectrometry. *Protein Sci.* **7**, 706–719 (1998).
- Claros, M. G. & Vincens, P. Computation method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* **241**, 779–786 (1996).
- Cedano, J., Aloy, P., Pérez-Pons, J. A. & Quero, E. Relation between amino acid composition and cellular location of proteins. *J. Mol. Biol.* **266**, 594–600 (1997).
- Boulikas, T. Nuclear localization signals (NLS). *Crit. Rev. Euk. Gene Exp.* **3**, 193–227 (1993).
- Li, P. *et al.* Cytochrome *c* and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade. *Cell* **91**, 479–489 (1997).
- Yasuhara, N. *et al.* Essential role of active nuclear transport in apoptosis. *Genes to Cells* **2**, 55–64 (1997).
- Enari, M. *et al.* A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* **391**, 43–50 (1998).
- Scaffidi, C. *et al.* Two CD95 (APO-1/Fas) signalling pathways. *EMBO J.* **17**, 1675–1687 (1998).
- Kluck, R. M. *et al.* Cytochrome *c* activation of CPP32-like proteolysis plays a critical role in a *Xenopus* cell-free apoptosis system. *EMBO J.* **16**, 4639–4649 (1997).
- Robinson, K. M. & Lemire, B. D. A requirement for matrix processing peptidase but not for mitochondrial chaperonin in the covalent attachment of FAD to yeast succinate dehydrogenase flavoprotein. *J. Biol. Chem.* **271**, 4061–4067 (1996).
- Liu, X., Zou, H., Slaughter, C. & Wang, X. DFF, a heterodimeric protein that functions downstream of caspase 3 to trigger DNA fragmentation during apoptosis. *Cell* **89**, 175–184 (1997).
- Samejima, K. *et al.* Transition from caspase-dependent to caspase-independent mechanisms at the onset of apoptotic execution. *J. Cell Biol.* **143**, 225–239 (1998).
- Oberhammer, F. *et al.* Apoptotic death in epithelial cells: cleavage of DNA to 300 and/or 50 kb fragments prior to or in the absence of internucleosomal fragmentation. *EMBO J.* **12**, 3679–3684 (1993).
- Lagarkova, M. A., Iarva, O. V. & Razin, S. V. Large-scale fragmentation of mammalian DNA in the course of apoptosis proceeds via excision of chromosomal DNA loops and their oligomers. *J. Biol. Chem.* **270**, 20239–20241 (1995).
- Trbovich, A. M. *et al.* High and low molecular weight DNA cleavage in ovarian granulosa cells: characterization and protease modulation in intact cells and in cell-free nuclear autodigestion assays. *Cell Death Differ.* **5**, 38–49 (1998).
- Susin, S. A. *et al.* The central executioner of apoptosis. Multiple links between protease activation and mitochondria in Fas/Apo-1/CD95- and ceramide-induced apoptosis. *J. Exp. Med.* **186**, 25–37 (1997).
- Bossy-Wetzel, E., Newmeyer, D. D. & Green, D. R. Mitochondrial cytochrome *c* release in apoptosis

- occurs upstream of DEVD-specific caspase activation and independently of mitochondrial transmembrane depolarization. *EMBO J.* 17, 37–49 (1998).
27. Marzo, I. *et al.* Bax and adenine nucleotide translocator cooperate in the mitochondrial control of apoptosis. *Science* 281, 2027–2031 (1998).
28. Enari, M., Hase, A. & Nagata, S. Apoptosis by a cytosolic extract from Fas-activated cells. *EMBO J.* 14, 5201–5208 (1995).
29. Wada, J. & Kanwar, Y. S. Characterization of mammalian translocase of inner mitochondrial membrane (Tim44) isolated from diabetic newborn mouse kidney. *Proc. Natl Acad. Sci. USA* 95, 144–149 (1998).
30. Zhu, W. *et al.* Bcl-2 mutants with restricted subcellular localization reveal spatially distinct pathways for apoptosis in different cell types. *EMBO J.* 15, 4130–4141 (1996).

Acknowledgements. We thank P. M. Alzari for suggestions, M. Geuskens for electron microscopy, G. Salvesen for caspase-8, D. Andrews for Rat-1 cells, and S. Arya and S. Chung for expression constructs. This work was supported by grants from ANRS, ARC, CNRS, FE, FRM, INSERM, LNC, and PRFMIP (to G.K.), the NIH, the NSF Science and Technology Center for Molecular Biotechnology (to R.A.), and Amgen (to D.P.S. and J.M.P.). S.A.S. and I.M. hold postdoctoral fellowships from the European Commission and from the Spanish Ministry of Science, respectively.

Correspondence and requests for materials should be addressed to G.K. (e-mail: kroemer@infobiogen.fr).

The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1

Soojin Ryu, Sharleen Zhou, Andreas G. Ladurner & Robert Tjian

Howard Hughes Medical Institute, University of California at Berkeley, Molecular and Cell Biology, 401 Barker Hall, Berkeley, California 94720-3204, USA

Activation of gene transcription in metazoans is a multistep process that is triggered by factors that recognize transcriptional enhancer sites in DNA. These factors work with co-activators to direct transcriptional initiation by the RNA polymerase II apparatus¹. One class of co-activator, the TAF_{II} subunits of transcription factor TFIID, can serve as targets of activators and as proteins that recognize core promoter sequences necessary for transcription initiation^{2–5}. Transcriptional activation by enhancer-binding factors such as Sp1 (ref. 6) requires TFIID, but the identity of other necessary cofactors has remained unknown. Here we describe a new human factor, CRSP, that is required together with the TAF_{II}s for transcriptional activation by Sp1. Purification of CRSP identifies a complex of approximate relative molecular mass 700,000 ($M_r \sim 700K$) that contains nine subunits with M_r values ranging from 33K to 200K. Cloning of genes encoding CRSP subunits reveals that CRSP33 is a homologue of the yeast mediator subunit Med7 (ref. 7), whereas CRSP150 contains a domain conserved in yeast mediator subunit Rgr1 (ref. 8). CRSP p200 is identical to the nuclear hormone-receptor co-activator subunit TRIP2/PBP^{9,10}. CRSPs 34, 77 and 130 are new proteins, but the amino terminus of CRSP70 is homologous to elongation factor TFIIS¹¹. Immunodepletion studies confirm that these subunits have an essential cofactor function. The presence of common subunits in distinct cofactor complexes suggests a combinatorial mechanism of co-activator assembly during transcriptional activation.

Sp1 is important in the transcriptional regulation of genes from *Drosophila* to humans. Previous studies showed that the interaction of the glutamine-rich activation domains of Sp1 with the TAF_{II} subunits of TFIID (where TAF is TATA-binding protein (TBP)-associated factor) is required for *in vitro* transcriptional activation⁴. However, these early studies were done using a partially purified transcription system and did not address the question of whether another cofactor(s) may be necessary to potentiate activation by Sp1 fully. Other cofactor fractions may indeed enhance activator-dependent transcription *in vitro*¹². One such crude fraction (USA)

is known to contain many positive and negative cofactors, including PC4, a single-stranded DNA-binding protein, and PC2, a complex of $M_r > 500K$, whose contribution to transcription activation by Sp1 has not been fully defined^{13–15}. We therefore set out to characterize the cofactors that are required for activation by Sp1.

To identify potential Sp1 cofactors, we first developed an *in vitro* transcription assay that consisted of purified recombinant TFIID, partially purified RNA polymerase II and basal transcription factors TFIIB, E, F and H, but contained no additional cofactor fraction such as USA. This mixture was supplemented with either antibody-affinity purified TFIID or recombinant TBP. As expected, reactions lacking TFIID or TBP (Fig. 1a) failed to support any detectable transcription from either the test template containing three GC-box Sp1-recognition sites or a control template lacking Sp1-binding sites. Thus, none of the partially purified general transcription factors is substantially contaminated with endogenous TFIID. When we added immunopurified human TFIID to the transcription system, we detected basal levels of transcription from both templates but little Sp1-directed activation (Fig. 1a). By contrast, when we added a phosphocellulose fraction (P1M) containing endogenous TFIID and several putative cofactor activities (including USA)

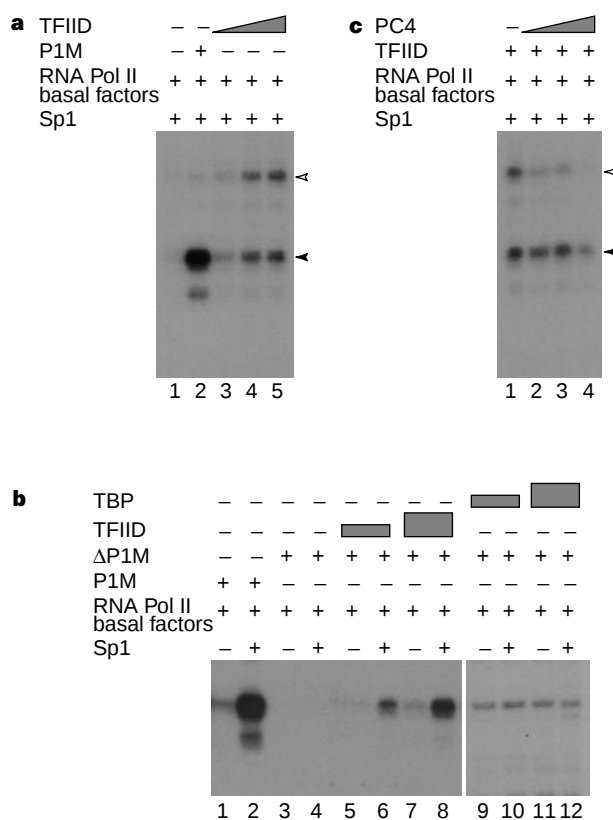


Figure 1 Transcriptional activation by Sp1 requires a new cofactor. **a**, Factor requirements for Sp1-dependent activation. Transcription reactions contain Sp1 (lanes 1–5), RNA polymerase II and basal transcription factors (lanes 1–5), the 1 M phosphocellulose fraction (P1M; lane 2) and increasing amounts of immunopurified TFIID (lanes 3–5). Transcription from a GC-box-containing template and a control DNA template lacking Sp1-binding sites is indicated by filled and open arrowheads, respectively. **b**, TFIID and a cofactor in the P1M fraction are needed to potentiate Sp1-dependent activation. Transcription reactions contain the GC-box template (lanes 1–12), Sp1 (lanes 2, 4, 6, 8, 10 and 12), RNA polymerase II and basal transcription factors (lanes 1–12), the P1M fraction (lanes 1 and 2), the P1M fraction treated with anti-TBP affinity resin to deplete TFIID (Δ P1M, lanes 3–12) and increasing amounts of immunopurified TFIID (lanes 5–8) or TBP (lanes 9–12). **c**, PC4 cannot substitute for the cofactor(s) present in the Δ P1M fraction. Transcription reactions were performed as described in **a**, in the presence of increasing amounts of PC4 (lanes 2–4) instead of the Δ P1M or P1M fraction.

to these reactions, we observed strong activation (20-fold) of transcription from the GC-box-containing template but not from the control template. The P1M fraction therefore contains one or more cofactors required for activation by Sp1.

To uncover new cofactor activities in the TFIID-containing P1M fraction, we had to remove contaminating TFIID. We efficiently depleted both TFIID and TBP from this fraction by multiple rounds of immunoprecipitation with anti-TBP antibodies. As expected, the resulting TFIID-free cofactor fraction no longer supported basal or activated transcription (Fig. 1b). The addition of immunopurified TFIID containing the full complement of TAF_{II} subunits together with the depleted cofactor fraction (Δ P1M) restored robust activation by Sp1 (Fig. 1b). Addition of purified recombinant TBP and the cofactor fraction restored basal transcription but not responsiveness to Sp1 (Fig. 1b). This indicates that efficient activation by Sp1 requires both TFIID and one or more cofactor activities present in the P1M fraction. Because the P1M fraction probably contains the positive cofactor PC4, we next tested the ability of purified recombinant PC4 to substitute for the P1M activity. PC4 inhibited basal transcription and was unable to support activation by Sp1 when added in combination with TFIID (Fig. 1c). At higher concentrations, PC4 appeared to inhibit transcription from both templates.

To characterize the polypeptide composition and biochemical properties of the new Sp1 cofactor, we purified the cofactor activity through five consecutive chromatographic columns and glycerol gradient sedimentation (Fig. 2a). After each chromatographic step, we identified the fractions containing the activity by assaying for Sp1-dependent activation (Fig. 2b) in the absence of added USA. We estimated that after MonoS anion-exchange chromatography,

the specific activity of the purified cofactor was ~50,000–100,000-fold higher than in crude nuclear extracts. The highly purified fraction thus contained a cofactor required for Sp1 activation (CRSP) that is distinct from PC4. As CRSP can replace the USA fraction, it may represent one of the activities present in USA, such as PC2. However, because the identity of the polypeptide(s) representing PC2 activity has not been defined, we cannot yet ascertain the relationship between CRSP and PC2. As a preliminary biochemical characterization of CRSP, we performed both glycerol gradient sedimentation and gel-filtration analysis to assess the native relative molecular mass of this cofactor (data not shown). Our results indicate that CRSP behaves as a single complex with an apparent M_r value of 700K, consistent with it being a multisubunit cofactor.

Using nuclear extracts prepared from 400 litres of HeLa cells, we obtained ~20 μ g of purified CRSP. SDS–polyacrylamide gel electrophoresis (SDS–PAGE) of the most extensively purified MonoS fractions revealed a pattern of nine to ten polypeptides that consistently co-purified with CRSP transcriptional activity (Fig. 2c, d). The polypeptides that co-chromatographed with CRSP included species of 33K, 34K, 70K, 77K, 130K and 150K. Most of the preparations also contained polypeptides of 85K, 100K and 200K. In contrast, the 175K band failed to co-purify on a reproducible basis. We therefore tentatively identified nine polypeptides as putative CRSP subunits and targeted them for further analysis.

The MonoS-purified CRSP was subjected to preparative SDS–PAGE separation, followed by transfer to nitrocellulose membrane and extraction by LysC protease digestion of individual polypeptides. The proteolytic peptides were purified by reversed-phase high-performance liquid chromatography (HPLC) and subjected

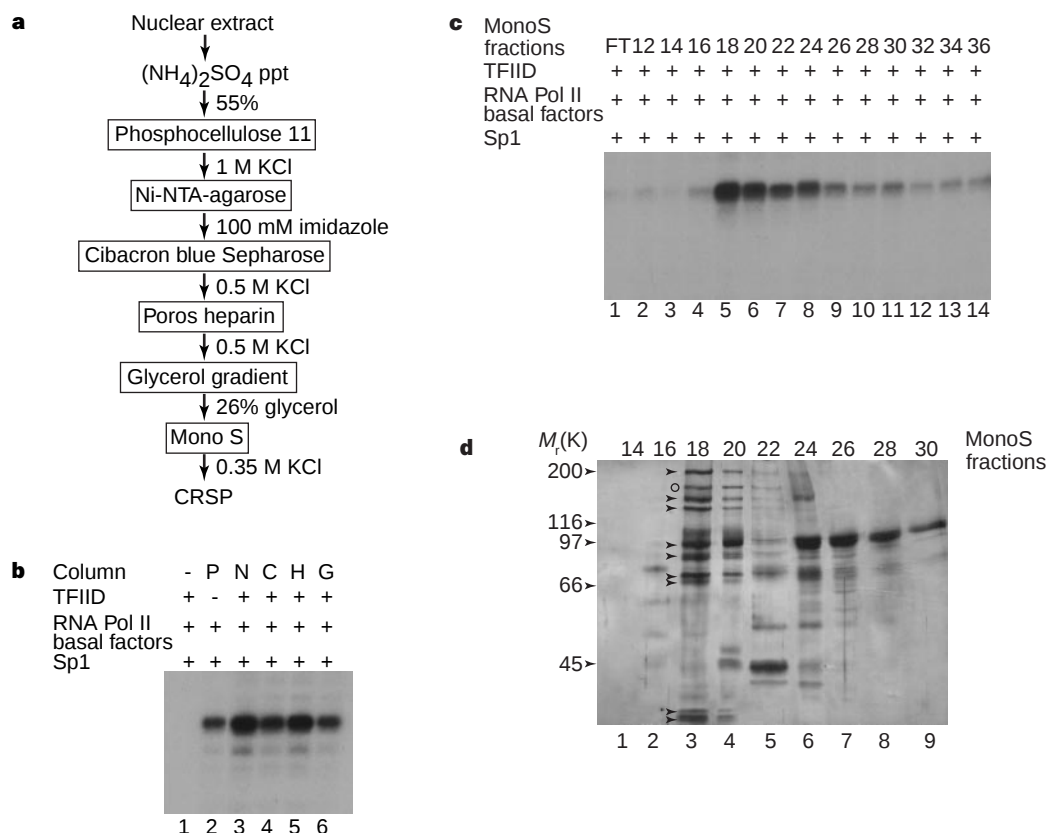


Figure 2 Purification of CRSP. **a**, Procedure for purification of CRSP. **b**, Reconstituted transcription reactions with the GC-box template, Sp1, RNA polymerase II and basal transcription factors (lanes 1–6) and immunopurified TFIID (lanes 1, 3–6) were supplemented with CRSP fractions recovered from phosphocellulose (P), Ni-NTA-agarose (N), Cibacron Blue Sepharose (C), Poros heparin (H) or glycerol gradient centrifugation (G). **c**, Transcription profile of CRSP

activity after MonoS chromatography. Reactions were supplemented with the indicated MonoS fractions (lanes 1–14) or flow-through (FT) fraction. **d**, Silver-stained SDS–PAGE gel of the MonoS fractions. Arrowheads indicate the proteins that co-migrate with CRSP activity. The open circle indicates a protein that fails to co-purify with CRSP in a reproducible manner.

to microsequence peptide analysis. Multiple peptide sequences derived from seven subunits were used to generate probes and/or to identify expressed sequence tag (EST) sequences. Several probes obtained by this strategy were used to screen human complementary DNA libraries for sequences encoding the putative CRSP subunits. (See Supplementary Information for the amino-acid sequences of the identified subunits of CRSP.)

The sequences show that CRSP represents a new cofactor complex for Sp1 that shares some subunits with other cofactors but also contains some unique subunits. CRSP34 and CRSP130 encode new human proteins with no apparent homology to other known gene products. In contrast, both CRSP33 and CRSP150 share homology to components of the yeast mediator, a complex that can potentiate activated transcription in yeast¹⁶. CRSP33 is a human homologue of yeast Med7 and the CRSP33 sequence is identical to the recently reported sequence of the human MED7 protein⁷. CRSP150 is identical to EXLM1, a novel protein of unknown function¹⁷. Residues 92–164 of CRSP150 show significant similarity to a yeast mediator component, Rgr1 (Fig. 3)⁸. The homology between the two proteins outside this region is not significant. In addition, CRSP77 and CRSP150 contain peptide sequences identical to those described for the p78 and p110 subunits of a murine complex that contains subunits similar to those of the yeast mediator¹⁸. CRSP150 contains peptide sequences identical to those described for the p150 subunit of NAT, a recently identified human complex that appears negatively to regulate activated transcription¹⁹. Although both Med7 and Rgr1 have previously been associated with a mediator complex, the other prototypic 'mediator' subunits, such as Srb proteins, Gal11 and the other Med proteins, do not appear to be present in CRSP. Instead, proteolytic peptides of the p200 subunit of CRSP match peptides of the TRIP2/PBP/RB18A protein^{9,10,20}, which has also been described, more recently, as DRIP230 and TRAP220, a component of a large co-activator complex that may be involved in inducible activation by nuclear hormone receptors such as peroxisome proliferator-associated receptor- γ , retinoid-X receptor, vitamin D receptor and others^{21,22}. In addition, the N-terminal domain of CRSP70 shows similarity to transcription elongation factor SII (Fig. 3)¹¹.

We used the cDNA clones encoding the CRSP subunits to generate partial or complete recombinant proteins. We then raised polyclonal antisera directed against CRSP subunits and used these antisera to characterize the biochemical and functional properties of CRSP. We attempted to generate antibodies against most of the CRSP subunits; however, only the antisera against CRSP33, CRSP130 and CRSP150 efficiently recognized denatured proteins on western blots. Furthermore, only anti-CRSP150 antibody was able to immunoprecipitate the endogenous complex. We therefore used antigen-affinity-purified anti-CRSP150 antibodies to determine which polypeptides form a stable CRSP complex and

to test whether these polypeptides are responsible for CRSP's cofactor activity. Anti-CRSP150 antibodies co-immunoprecipitated a complex of nine to ten polypeptides from the partially purified Ni-NTA-agarose chromatographic fraction. The polypeptide composition of the immunopurified complex was nearly identical to that of the most pure MonoS fractions, consisting of CRSPs 33, 34, 70, 77, 85, 100, 130, 150 and 200 (Fig. 4a). In addition, western blot analysis (see Supplementary Information) confirmed that the immunoprecipitated complex contained CRSPs 33, 130 and 150. Most of these subunits appeared to be stably associated into a complex that survived high-salt (1 M) washes (Fig. 4a, lane 4). We also confirmed that at least CRSPs 33, 130 and 150 co-purified as a stable

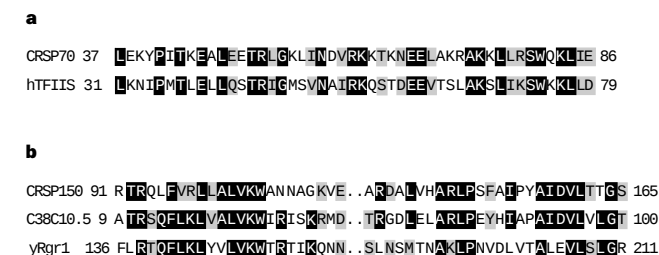


Figure 3 Sequence alignments between CRSP subunits and sequence homologues. **a**, The N terminus of CRSP70 shares sequence homology with the N terminus of the human transcription elongation factor IIS. **b**, The N-terminal region of CRSP150 shares significant sequence homology with the N-terminal region of the yeast mediator component Rgr1 and a novel *Caenorhabditis elegans* protein, C38C10.5. Black shading denotes identical residues; grey shading denotes similar residues.

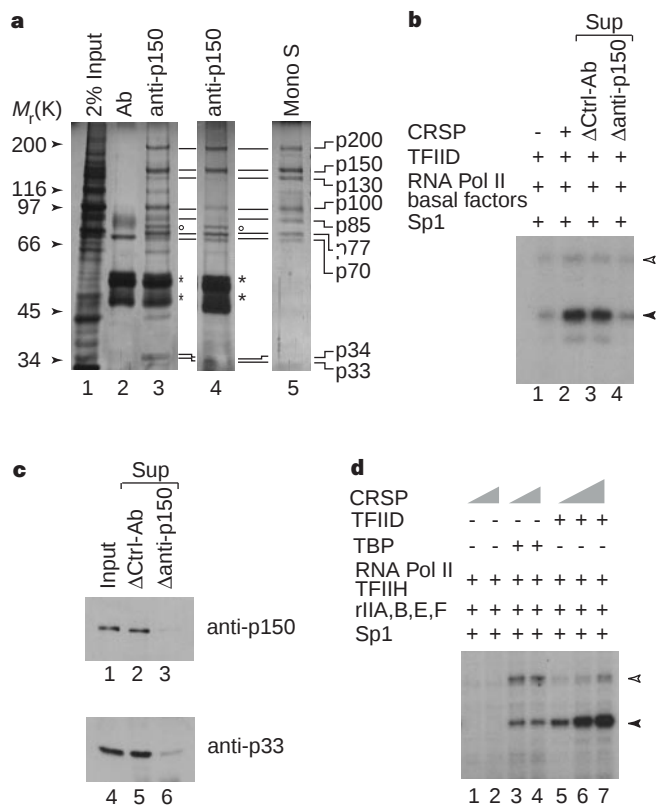


Figure 4 Immunoprecipitation of the CRSP complex and immunodepletion of CRSP activity. **a**, Comparison of polypeptides immunopurified with anti-p150 antibodies (lanes 3 and 4) to MonoS-purified CRSP (lane 5). Lanes 1 and 2 show the starting material for immunopurification and the antibody alone, respectively. The nine CRSP subunits are indicated by their respective apparent molecular weights. Asterisks denote the immunoglobulin heavy chain. Open circles indicate a protein that immunoprecipitated but is not present in the MonoS fraction. Lanes 3 and 4 show immunoprecipitation patterns after washes with 0.7 M and 1 M KCl, respectively. **b**, Immunodepletion of CRSP reduces transcription activated by Sp1. Transcription reactions were reconstituted as described (Fig. 1) and supplemented with a partially purified CRSP fraction (see Methods) (lane 2) or with the same fraction after immunodepletion with control antibodies (lane 3) or anti-p150 antibodies (lane 4). Transcription from a GC-box-containing template and a control template is indicated by the filled and open arrowheads, respectively. **c**, Western blots of the immunodepleted fractions produced using anti-p150 (lanes 1–3) and anti-p33 (lanes 4–6) antibodies. Lanes 1 and 4, the partially purified CRSP fraction (see **b**); lanes 2 and 5, supernatant of the same fraction after depletion with control antibodies; lanes 3 and 6, supernatant of the same fraction after depletion with anti-p150 antibodies. **d**, CRSP mediates Sp1-dependent activation in a highly purified transcription system. Transcription reactions were reconstituted with purified recombinant Sp1, recombinant TFIID, B, E and F (rIIA, B, E, F), RNA polymerase II, affinity-purified TFIID²³ and highly purified CRSP alone (lanes 1 and 2) or with CRSP in the presence of recombinant TBP (lanes 3 and 4) or in the presence of immunopurified TFIID (lanes 5–7).

subcomplex during glycerol gradient sedimentation and gel filtration (data not shown). Most important, immunodepletion of a highly purified CRSP fraction with anti-CRSP150 antibody markedly reduced cofactor activity (Fig. 4b, c). This result indicates that the complex co-precipitating with CRSP150 contributes to the transcriptional cofactor function of CRSP. We cannot exclude the possibility that other weakly associated polypeptides that are not found stably associated with CRSP150 may also contribute to CRSP activity.

So far, our identification and purification of CRSP relied largely on a reconstituted transcription system consisting of partially purified basal factors. As a step towards developing a fully defined transcription system to study activators and their co-activator requirements, we tested the activity of CRSP in a highly purified system. This consisted of recombinant TFIIA, B, E and F subunits, Sp1, affinity-purified TFIIF²³, conventionally purified RNA polymerase II, and highly purified CRSP, in the presence of recombinant TBP or immunopurified TFIID (Fig. 4d). Our results indicate that both CRSP and TFIID are required to mediate Sp1-dependent activation in this highly purified transcription system, thus establishing CRSP as an essential cofactor for Sp1-mediated transcription.

The subunit composition of CRSP indicates that distinct cofactor complexes contain common subunits. For example, CRSP contains three subunits that are also present in yeast or the putative murine mediator complex^{7,8,18}, CRSPs 33, 77 and 150, but not other mediator subunits such as Srb proteins, Med proteins and Gal11 (ref. 16). Yeast and murine mediator, on the other hand, do not appear to contain homologues of CRSP70, CRSP130 and the p200 subunit of CRSP. The CRSP/mediator comparison is reminiscent of the recently uncovered relationship between TFIID and the SAGA/PCAF complex required for transcriptional stimulation: some of the TAF_{II} subunits of TFIID are also integral components of the SAGA/PCAF complex²⁴. In addition, CRSP shares subunits with NAT¹⁹ (CRSP150) and the DRIP/TRAP co-activator^{21,22} (p200 subunit of CRSP). As the sequences for many of the NAT-, TRAP- and DRIP-complex subunits have not yet been published, it is possible that there may be other shared subunits. However, both TRAP and DRIP are larger complexes containing a number of subunits with molecular masses that are not found in CRSP.

An increasing number of molecules with co-activator functions, such as the TAF_{II}s, CBP, SRC, PC4, SRBs and TRAP/DRIP subunits, have been identified^{5,14,21,25–27}. It was, however, unclear which mechanisms might have evolved to generate the diversity and specificity of co-activators necessary to accommodate the large number of gene-specific promoters and their attendant activators in animal cells. Our results indicate that one way in which co-activator complexes with distinct activities could be assembled could be the mixing and matching of co-activator subunits or subcomplexes in a combinatorial fashion. The finding that CRSP, a co-activator complex required for Sp1 activation, contains several potentially unique subunits, as well as some subunits in common with the yeast and mouse mediators, DRIP/TRAP complexes and the NAT complex, supports the principle that distinct composite co-activator complexes may be assembled by a combinatorial mechanism. As part of such a model, we envisage that different co-activator complexes are directed to assemble at specific promoters by virtue of their interactions with activators, with each other, with core promoter sequences and with the basal machinery. An alternative model would invoke a large collection of partially or completely pre-assembled cofactors. It will be important to determine how CRSP functions in other contexts. Does it function with other activators and, if so, how do different activators use the same cofactor? □

Methods

In vitro transcription and preparation of transcription factors. *In vitro* transcription assays were done with a HeLa fractionated system as described²⁸, with few modifications. The P1M fraction was prepared by loading a nuclear

extract on a P11 (Whatman) column at 0.3 M KCl concentration, washing the extract with 0.5 M KCl and eluting it at 1 M KCl. For the RNA polymerase II basal factor fractions, the extract was dialysed to 20 mM KCl, applied to a P11 column at 0.1 M KCl, washed at 0.3 M KCl and eluted at 0.5 M KCl. Sp1 was overexpressed in HeLa cells using a vaccinia-virus-expressing vector and purified to near homogeneity as described⁴. Immunopurified TFIID was prepared from the P1M fraction using a monoclonal antibody against hTAF_{II}130 immobilized on protein A-Sepharose resin (Pharmacia) and eluted with a 20-mer peptide containing the epitope. dTFIIA was overexpressed in *Escherichia coli* and prepared as described²⁹. Transcription assays were performed as described². The DNA templates used included (GC)₃ BCAT³, which contains three Sp1-binding sites upstream of the E1B TATA box, and +15 BCAT, which lacks Sp1-binding sites and contains a 15-base-pair insertion producing a longer transcript. Typical reactions (25 µl) contained 50 ng of each template, 5 ng dTFIIA, 50 ng Sp1, 300 ng RNA polymerase II basal factor fraction, and 5–25 ng immunopurified TFIID, or 10 ng recombinant TBP plus varying amounts of CRSP fractions. For the experiment shown in Fig. 1c, 5–200 ng of recombinant PC4 was used¹³. For transcription assays in the purified system, recombinant hTFIIB, F and E and RNA polymerase II fractions were prepared as described³⁰. TFIIF was affinity-purified as described²³. Reactions contained 50 ng dTFIIA, 50 ng Sp1, 5 ng TFIIB, 50 ng TFIIE34, 50 ng TFIIE56, 80 ng TFIIF, 1 ng immunopurified TFIIF, 5 ng immunopurified TFIID, 5 ng RNA polymerase II and 50 ng of a highly purified MonoS CRSP fraction. The reaction products were detected by primer extension. Each transcription reaction was repeated several times, and representative data are shown.

Purification of CRSP. The P1M fraction was used as starting material and HEMG (25 mM HEPES, 12.5 mM MgCl₂, 0.2 mM EDTA and 10% glycerol, pH 7.6) supplemented with KCl was used for all of the purification steps, except during Ni-NTA-agarose purification, when EDTA was omitted. The CRSP-containing P1M fraction was loaded on a Ni-NTA-agarose resin (Qiagen) in the presence of 5 mM imidazole, 0.1% N-P40 and 5 mM β-mercaptoethanol, washed with 25 mM HEPES, 12.5 mM MgCl₂, 0.7 M KCl and 10 mM imidazole, and eluted with the same buffer containing 100 mM imidazole and 0.15 mM KCl. The Ni-NTA fractions containing CRSP activity were loaded directly onto a Cibacron Blue Sepharose column (Pharmacia), washed with HEMG/0.2 M KCl and eluted with HEMG/0.5 M KCl. The protein peaks were pooled, dialysed to HEMG/0.2 M KCl, applied to Poros heparin (Perceptive Biosystems) and eluted with a gradient from 0.2 M to 1 M KCl. Heparin fractions containing CRSP activity (0.5 M) were pooled and applied to a 15–30% glycerol gradient (in HEMG/0.2 M KCl) and centrifuged at 50,000 r.p.m. for 7 h (TLS55 rotor, Beckman). The active fractions were pooled and applied to MonoS PC 1.6/5 (Pharmacia) and eluted with a gradient from 0.2 M to 1 M KCl. Further details of the purification will be published elsewhere.

Cloning of cDNAs encoding CRSP subunits. MonoS fractions obtained from 400 litres of HeLa cells were resolved by SDS-PAGE, transferred to nitrocellulose membrane, and stained with Ponceau S (Sigma). The bands corresponding to the CRSP subunits were excised and digested with LysC protease. Peptides eluted from the membrane were resolved by reversed-phase HPLC and microsequenced. The amino-acid sequences of peptides of the identified seven CRSP subunits (see Supplementary Information) were used to determine corresponding EST sequences using the BLAST program^{7,21,22}. A partial clone of CRSP150 was identical to the EXLM1 gene (GenBank accession number, AB006651). Full-length CRSP34 was assembled from EST clones 147,371 and 182,652. CRSP77 was assembled using human EST clone 262,073 and those clones represented in the THC 192416 alignment (Institute for Genome Research). For CRSP130 and CRSP70 sequences, cDNAs obtained from EST clones 767,814 and 332,225, respectively, were used to probe a human testis library (Clontech). Positive clones were cloned into the pBluescript KS⁺ plasmid (Stratagene), sequenced and aligned.

Antibody generation and immunoprecipitation. Residues 1,306–1,454 of CRSP150, residues 366–500 of partial CRSP130, and residues 51–167 of CRSP33 were overexpressed as either His₆ tag (Invitrogen) or glutathione-S-transferase (GST; Pharmacia) fusion proteins in *E. coli*. Rabbits were boosted five to seven times with affinity-purified antigen. The antisera obtained were affinity-purified using antigens immobilized on Affigel 10 or 11 resin (Biorad). For the immunoprecipitation assay shown in Fig. 1a, 100–200 ng affinity-purified

anti-CRSP150 antibody was incubated with 100 µg of CRSP fractions derived from the Ni-NTA-agarose chromatography. The immunoprecipitated complex was recovered by addition of protein A-Sepharose, washed extensively with HEMG/0.7 M KCl or HEMG/1 M KCl and extracted with 100 mM glycine buffer, pH 2.5. For the immunodepletion studies shown in Fig. 4b, c, the CRSP fraction obtained after Cibacron Blue Sepharose chromatography was concentrated over a MonoS column. 200 ng of eluted material was incubated with 100 ng affinity-purified anti-CRSP150 antibody or control antibody (unrelated rabbit serum IgG). Antibody-antigen complexes were recovered using protein A-Sepharose and the supernatant was assayed for CRSP activity in transcription assays.

Received 22 October; accepted 9 December 1998.

1. Tjian, R. & Maniatis, T. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* **77**, 5–8 (1994).
2. Pugh, B. F. & Tjian, R. Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* **61**, 1187–1197 (1990).
3. Dynlacht, B. D., Hoey, T. & Tjian, R. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* **66**, 563–576 (1991).
4. Hoey, T. *et al.* Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties expected of coactivators. *Cell* **72**, 247–260 (1993).
5. Mengus, G. *et al.* Cloning and characterization of hTAFII18, hTAFII20 and hTAFII28: three subunits of the human transcription factor TFIID. *EMBO J.* **14**, 1520–1531 (1995).
6. Kadonaga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**, 1079–1090 (1987).
7. Myers, L. C. *et al.* The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **12**, 45–54 (1998).
8. Li, Y. *et al.* Yeast global transcriptional regulators Sin4 and Rgr1 are components of mediator complex/ RNA polymerase II holoenzyme. *Proc. Natl Acad. Sci. USA* **92**, 10864–10868 (1995).
9. Lee, J. W., Choi, H. S., Gyuris, J., Brent, R. & Moore, D. D. Two classes of proteins dependent on either the presence or absence of thyroid hormone for interaction with the thyroid hormone receptor. *Mol. Endocrinol.* **9**, 243–254 (1995).
10. Zhu, Y., Qi, C., Jain, S., Rao, M. S. & Reddy, J. K. Isolation and characterization of PBP, a protein that interacts with peroxisome proliferator-activated receptor. *J. Biol. Chem.* **272**, 25500–25506 (1997).
11. Sekimizu, K., Kobayashi, N., Mizuno, D. & Natori, S. Purification of a factor from Ehrlich ascites tumor cells specifically stimulating RNA polymerase II. *Biochemistry* **15**, 5064–5070 (1976).
12. Meisterernst, M., Roy, A. L., Lieu, H. M. & Roeder, R. G. Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. *Cell* **66**, 981–993 (1991).
13. Ge, H. & Roeder, R. G. Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell* **78**, 513–523 (1994).
14. Kretschmar, M., Stelzer, G., Roeder, R. G. & Meisterernst, M. RNA polymerase II cofactor PC2 facilitates activation of transcription by GAL4-AH *in vitro*. *Mol. Cell. Biol.* **14**, 3927–3937 (1994).

15. Brandsen, J. *et al.* C-terminal domain of transcription cofactor PC4 reveals dimeric ssDNA binding site. *Nature Struct. Biol.* **4**, 900–903 (1997).
16. Kim, Y. J., Björklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599–608 (1994).
17. Yoshikawa, H., Fujiyama, A., Nakai, K., Inazawa, J. & Matsubara, K. Detection and isolation of a novel human gene located on Xp11.2-p11.4 that escapes X-inactivation using a two-dimensional DNA mapping method. *Genomics* **49**, 237–246 (1998).
18. Jiang, Y. W. *et al.* Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc. Natl Acad. Sci. USA* **95**, 8538–8543 (1998).
19. Sun, X. *et al.* NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. *Mol. Cell* **2**, 213–222 (1998).
20. Drané, P., Barel, M., Balbo, M. & Frade, R. Identification of RB18A, a 205 kDa new p53 regulatory protein which shares antigenic and functional properties with p53. *Oncogene* **15**, 3013–3024 (1997).
21. Rachez, C. *et al.* A novel protein complex that interacts with the vitamin D3 receptor in a ligand-dependent manner and enhances VDR transactivation in a cell-free system. *Genes Dev.* **12**, 1787–1800 (1998).
22. Yuan, C. X., Ito, M., Fondell, J. D., Fu, Z. Y. & Roeder, R. G. The TRAP220 component of a thyroid hormone receptor-associated protein (TRAP) coactivator complex interacts directly with nuclear receptors in a ligand-dependent fashion. *Proc. Natl Acad. Sci. USA* **95**, 7939–7944 (1998).
23. LeRoy, G., Drapkin, R., Weis, L. & Reinberg, D. Immunoaffinity purification of the human multi-subunit transcription factor IIH. *J. Biol. Chem.* **273**, 7134–7140 (1998).
24. Struhl, K. & Moqtaderi, Z. The TAFs in the HAT. *Cell* **94**, 1–4 (1998).
25. Chrivia, J. C. *et al.* Phosphorylated CREB binds specifically to the nuclear protein CBP. *Nature* **365**, 855–859 (1993).
26. Onate, S. A., Tsai, S. Y., Tsai, M. J. & O'Malley, B. W. Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science* **270**, 1354–1357 (1995).
27. Kaiser, K. & Meisterernst, M. The human general cofactors. *Trends Biochem. Sci.* **21**, 342–345 (1996).
28. Dignam, J. D., Martin, P. L., Shastri, B. S. & Roeder, R. G. Eukaryotic gene transcription with purified components. *Methods Enzymol.* **101**, 582–598 (1983).
29. Yokomori, K. *et al.* *Drosophila* TFIIA directs cooperative DNA binding with TBP and mediates transcriptional activation. *Genes Dev.* **8**, 2313–2323 (1994).
30. Näär, A. M. *et al.* Chromatin, TAFs and a novel multiprotein coactivator are required for synergistic activation by Sp1 and SREBP *in vitro*. *Genes Dev.* **12**, 3020–3031 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank R. Bell for the preparation of HeLa cell nuclei, M. Haggart for DNA sequencing and DNA oligonucleotide synthesis; H. Yoshikawa for plasmids containing the EXLM1 gene; A. Näär, P. Beaurang, C. Inouye and B. D. Lemon for the components of the highly purified transcription system; D. Reinberg for the gift of TFIID antibody; and D. Rio, Q. Zhou, K. Collins, M. Holmes, R. Freiman, M. D. Rabenstein and B. D. Lemon for comments on the manuscript. S.R. is a Howard Hughes Medical Institute predoctoral fellow and A.G.L. is a Wellcome Trust International Prize Travelling Research Fellow. This work is supported in part by a grant from the NIH to R.T.

Correspondence and requests for materials should be addressed to R.T. Sequences for the cloned CRSP-subunit genes have been deposited with GenBank under accession numbers AF104251–56.

Bottles, washers and the like

Under the heading 'Biotechnology', twenty-four items that just might be what you are looking for to make life that little bit easier. Web links – hyperlinked on nature.com – are given where available.

Monoclonal antibodies

From TSD BioServices www.tsdbioservices.com

Mouse monoclonals, produced in vitro

Monoclonal antibodies devoid of extraneous murine can be obtained from TSD BioServices, based in Upper Saddle River, New Jersey. Their *in vitro* contract manufacturing service makes use of hollow-fibre bioreactors and roller bottle systems to produce gram quantities of monoclonal antibody routinely. Critical metabolic and production parameters are monitored and reported on the Certificate of Analysis at completion of the project. TSD BioServices are found on 518-537-800.

Reader Service No. 100

PCR tube storage

From Molecular BioProducts

www.mbpinc.com

A new packaging design to keep PCR tubes contamination-free

Instead of packaging PCR tubes in bulk bags, Molecular BioProducts have distributed them between smaller, individually ealed PurePaks. This maintains the integrity of the tubes, improving the reliability of PCR. Each PurePak is RNase-, DNase-, DNA- and pyrogen-free, and is available with the company's full line of PCR tubes, strips and plates.

Reader Service No. 101

Biofuge

From Kendro

www.kendro.com

Cold comfort for centrifuges

The Biofuge primo R centrifuge is now available in refrigerated form, as the Biofuge primo R. Vulnerable samples can be refrigerated at 4°C, even at maximum speed. Users can choose between five rotors. The 'highconic' fixed-angle rotor takes conical tubes with an angle of 45°, giving sedimentation at the bottom of the tube rather than the edge of the tube. Maximum acceleration is 10,000g. The high-performance microlitre rotor accelerates to 22,000g, recommended for DNA precipitation.

Reader Service No. 102

Effectene

Qiagen

www.qiagen.com

Transfection, economical with the DNA

Effectene Transfection Reagent is a non-liposomal lipid formulation offering potential advantages over many liposome reagents

and other transfection methods. Qiagen claim that Effectene Reagent, used together with a DNA-condensing enhancer, achieves exceptionally high transfection efficiencies with a wide variety of cell types, particularly with primary cells. The reagent is less toxic than many liposome reagents, and enables transfection in the presence of serum. In addition, Effectene Reagent delivers plasmid DNA into cells with a high efficiency, giving high transfection levels with relatively low amounts of DNA.

Reader Service No. 103

UpScale

From Pall Corporation

www.pall.com

Fully scalable filtration products for the pharmaceutical and biotechnology industries

The UpScale program is intended to simplify scale up from R&D pilot-scale and full production with a comprehensive line of filtration products. UpScale is a combined program from Pall BioPharmaceuticals and Pall Gelman Laboratory, which specialize, respectively, in process filtration/bio-separations and laboratory filtration. Pall products comprise a range of liquid service membranes including Ultipor (Nylon), Posidyne (positively charged Nylon), Supor (PES) and Fluorodyne II PVDF.

Reader Service No. 104

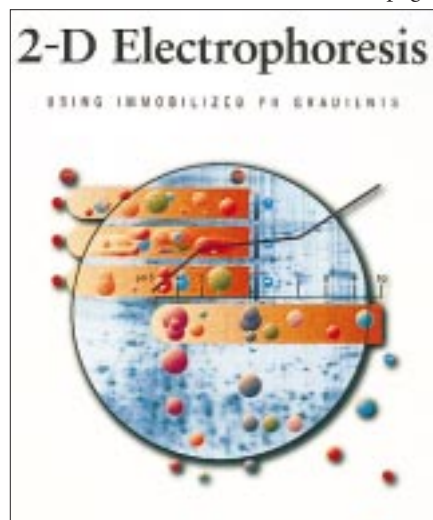
Electrophoresis manual

From Amersham Pharmacia

www.apbiotech.com

A technical guide to 2-D electrophoresis

Now available from Amersham Pharmacia Biotech, this manual boasts in its 56 pages



PAGE turner: electrophoresis tips, spirally bound.

an illustrated trouble-shooting guide with solutions to real gel experiments and the common problems associated with sample preparation and loading. There are four sections covering sample preparation, first-dimension IEF, second-dimension SDS PAGE and visualization and analysis of results.

Reader Service No. 105

Voyager

From Invitrogen

www.invitrogen.com

Recombinantly expressed protein in 100 per cent of cultured cells

Voyager Technology utilizes the VP22 gene product to target recombinant protein into virtually 100% of a culture cell population. The Voyager vectors, pVP22/myc-His and pVP22/myc-His TOPO, allow you to express your favourite gene as a fusion to VP22. Following transfection, the VP22 gene fusion is expressed and translocated to the nucleus of virtually 100% of the non-transfected cultured cells. Voyager vectors are suitable for the study of transcriptional and translational regulation as well as other nuclear regulation events.

Reader Service No. 106

AviTag

From Avidity

www.avidity.com

Enzyme for the biotinylation of peptide tags

AviTag is a 15-mer peptide that can be attached at the N- or C-terminus of any cloned protein and is recognized and biotinylated by biotin ligase (BirA enzyme). Using the AviTag peptide/BirA system, almost any protein can be specifically labelled with biotin. In conjunction with avidin or streptavidin, the AviTag system is now being widely used for the immobilization, purification and oligomerization of proteins. Avidity's BirA enzyme is supplied with a reaction mix containing all the reagents necessary to achieve >95% *in vitro* biotinylation. The enzyme is supplied essentially free of proteases.

Reader Service No. 107

Cloncapture

From Clontech

www.clontech.co.uk

A cDNA selection kit for the rapid capture of full-length cDNA clones

ClonCapture is a gene isolation system designed to enrich target clones from any plasmid library without extensive screening. Clontech claim that, using ClonCapture and

a partial cDNA probe, full-length target cDNA clones can be isolated in just 2 to 3 days, saving significant time compared with other methods. The protocol involves the ClonCapture RecA protein, a 200–600 bp biotinylated probe. The probe binds homologous cDNA clones which are captured by streptavidin-coated magnetic beads.

Reader Service No. 108

Vectamidine

From Accurate Chemical

www.accuratechemical.com

Transfection in 2 hours

Vectamidine (di-C14-tBut) is marketed as a rapid and easy-to-use transfection system reaching maximum activity in 2 hours. Vectamidine can function in the presence of serum, permits introduction of DNA and mRNA, and has a high response on adherent, non-adherent and primary culture cells. The reagent is stable for over one year or for 6 months in solution.

Reader Service No. 109

TnT Quick

From Promega

www.promega.com

Quick-fire transcription and translation

The TnT SP6 Quick coupled transcription/translation system produces proteins *in vitro* directly from DNA encoding a protein downstream of an SP6 RNA polymerase promoter. TnT Quick Master Mix includes all the components necessary for *in vitro* translation. TnT SP6 Quick Coupled Transcription/Translation System can be used in protein:protein interaction, protein:DNA/RNA interactions (gel shifts), rapid screening of mutants, rapid confirmation of gene products, simultaneous expression of multiple gene products, *in vitro* expression cloning/functional genomics and in protein activity studies.

Reader Service No. 110

Laboratory mice

From Taconic Biotechnology

www.taconic.com

Tumour-carrying rodents for chemotherapeutic and tumour biology research

Laboratory rats and mice implanted with specific tumour cell lines are available from Taconic Biotechnology. Services include inoculation to histocompatible hosts and xenographs to immunodeficient mice. Because they are effective in accepting non-murine tumour cells, athymic nude mice and severe combined immune deficient (SCID) mice are amongst the most frequently used hosts for tumour production. Protocols are available for the production of tumour bearing animals, diet manipulations and dosing before and after inoculation.

Reader Service No. 111



Keep it dark: the FP-2500 is self-contained.

FP-2500 gel visualizer

From Ultra-Lum

www.ultralum.com

Documentary evidence of gels

The FP-2500 is a gel visualization/documentation system designed to perform complete visualization and documentation of gels, stains or autoradiography results including oversized samples up to 20 × 40 cm in size in a benchtop darkroom cabinet. It includes a darkroom cabinet with viewing window, high-resolution CCD camera, motorized zoom lens, printer and monochrome monitor.

Reader Service No. 112

AccuTrac

From Commonwealth Biotechnologies

www.cbi-biotech.com

An aid to automated sequencing of DNA

AccuTrac, announced in January of this year, is a loading dye used to identify individual lanes in DNA samples. The AccuTrac process uses a fluorescent dye label that marks every lane on an acrylamide gel, which holds the samples that are read by the automated sequencer. All lanes are identified whether they contain data or not. The idea is that by preventing the possible misidentifications caused by "skipped" lanes, AccuTrac removes the need for manual review and lane assignment.

Reader Service No. 113

NicKit

From Vector Laboratories www.vectorlabs.com

Probe size optimization

NicKit *p.s.o.*, a new probe size optimization kit, has been developed for reducing the size of DNA probes to an optimal length for *in situ* hybridization, without the problem of over- or under-digestion encountered with other methods. The procedure involves a

brief UV irradiation of probe DNA using a transilluminator or hand-held UV lamp followed by enzymatic digestion using a specially prepared endonuclease. Phosphodiester bonds at pyrimidine dimers formed by UV irradiation are cleaved, resulting in average DNA lengths of 400 bases or less. Probe DNA can also react with NicKit *p.s.o.* reagents prior to labelling with PHOTO-PROBE Biotin, the FastTag system, or other non-destructive labelling methods. The NicKit *p.s.o.* reagents treat up to 300 µg of DNA.

Reader Service No. 114

Multichannel PMT

From Hamamatsu Photonics

www.hamamatsu.com

A multichannel photomultiplier tube for DNA sequencing

The R5319 Multichannel PMT has been developed especially for use with microtitre plates. Pitch and size of the 8 linear PMT channels match those of the plates to facilitate the direct measurement of optical signals. At one end of each channel is an independent photoelectrode giving a spectral response from 300 to 650 nm, and at the other end secondary electrons are collected by independent anodes. Suggested applications include fluorescence analysis, radioimmunoassay, enzyme immunoassay and chemiluminescence.

Reader Service No. 115

MACSelect

From Miltenyi Biotec

An alternative to establishing stably infected cells

MACSelect kits are designed to enrich transiently transfected cells, particularly in primary cells or other cell lines with a low number of transfected cells. For enrichment the cells are co-transfected with a pMACS vector encoding a truncated surface marker such as human CD4 or mouse H-2K^k, and the gene being studied. The kits are compatible with all commonly used transfection methods and Miltenyi make the claim that the system will make it possible to study gene expression in many formerly 'untransfectable' cells.

Reader Service No. 116

Filtration systems

From Millipore

www.millipore.com

Filtration for the pharmaceuticals industry

Millipore's Pellicon XL and the Labscale tangential flow filtration system are now available in North America through Fisher Scientific. These systems perform concentration, desalting, buffer exchange and clarification and are aimed particularly at the drug development market. Pellicon XL

processes samples for 50 to 1,000 ml using the same geometry as the larger Pellicon 2 for scalability. The Pellicon XL can run with the Labscale flow filtration system, docking directly onto the side of the reservoir.

Reader Service No. 117

Galaxy CO₂ incubator

From RS Biotech

Three incubators in one

For applications where frequent entry into the incubator is unavoidable, the Galaxy CO₂ incubator is available with three separate inner doors made of glass. These replace the standard inner door and effectively create three culture chambers so only one door need be opened at once. High-temperature decontamination, involving heating to 121 °C for 4 hours, is also an option, for elimination of bacteria and viruses.

Reader Service No. 118

pWEB cloning kit

From Cambio

www.cambio.co.uk

A cosmid cloning kit for blunt-end cloning of DNA fragments

With the pWEB kit it should be possible to create a cosmid library in two days. The kit utilizes a vector containing a blunt-ended cloning site (*Sma*I). First step in the process is shearing the genomic DNA by passage through a fine-gauge needle, end-repair using the enzyme mix and buffer, marking with T7 DNA and then running on a gel. The DNA is then extracted with Gelase, ligated into the cosmid, packaged with MAXPlax and transfected into host cells.

Reader Service No. 119

DNA synthesis

From Biosource

www.lifescrreen.com

DNA synthesis on tap

The Biosource plant at Keystone, USA, is taking orders via Appligene Oncor for DNA synthesis service, especially developed for complex syntheses such as long oligos, conjugations and FRET (or 'taqman') probes. The company works to a turn-around time of 1 week. What is more, for a limited period, the company is offering a free 'Art of Science' mousemat to anyone completing their DNA oligo questionnaire.

Reader Service No. 120

Aerosol spray bottle

From Nalgene

www.nalgene.com

A spray using no compressed air or propellants

The Nalgene air spray bottle is now available in a larger size, 180 ml capacity. It is made of translucent polypropylene, with polythene cap and rubber gasket, and is resistant to a wide variety of aqueous reagents. The aerosol is propelled by compressed air and



Pump it up: Nalgene's bottle, that is.

the bottle recharged by hand pumping.

Reader Service No. 121

EMBLA 1536

From Skatron

www.skatron.com

Microplate washing, in bulk

Skatron Instruments AS, based in Lier, Norway, produce a microplate washer for

the 384-well format, and included Windows based software and robotic integration into the package. The EMBLA 1536 Washer goes a step further, bringing the same technology to a 1,536-plate format.

Reader Service No. 122

Columbus Microplate Washer

From Tecan

www.tecan.com

A fully automated microplate washer

The Columbus offers a selection of 8, 12 or 16 channel manifold and allows use of up to four different wash buffers. Aspiration and dispensing are not pressurized so that any bottle type can be used for liquids without having air-leakage problems. For gentle washing of cell assays, the Columbus can add buffer in drips rather than as a jet, and either overflow or bottom washing can be performed as required by the user's ELISA kit. The Columbus is programmed by easy-to-use Windows-based software or via a built-in keyboard and display.

Reader Service No. 123

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

Definit residual DNA with Wako's
DNA Extractor Kit
 or **Isolate Genomic DNA with**
DNA Extractor WB Kit
 Both use the **Nal Method**
No Phenol No Chloroform

Wako

Wako BioProducts
 Wako Chemicals USA, Inc.
 1800 Redwood Road
 Richmond, VA 23137
 Telephone: (804) 871-7677
 Fax: (804) 871-7761
 (800) 992-6266
Wakoproducts@wako.com

READER ENQUIRY NO. 39

BMA BIOMEDICALS AG
 Your source for unique antibodies:
 MCP-1
 Scavenger receptor CD163
 Endothelial and Macrophage markers
 hMRP8/14 (Calprotectin) ELISA
 hIL-1 receptor antagonist ELISA

Call for free catalogue:
 ++41 61 811 6222 (phone)
 ++41 61 811 6006 (fax)
Info@bma.ch

READER ENQUIRY NO. 1

fluorometer.committed™

fluorometer.com
 A website of fluorometers and their
 application methods.

TURNER DESIGNS
 845 W. Maude Avenue • Sunnyvale, CA 94086
 (408) 749-0994 • FAX (408) 749-0998

READER ENQUIRY NO. 3

NEWCASTLE UNIVERSITY
 MOLECULAR BIOLOGY UNIT

PROTEIN SEQUENCING
Membrane-bound, solid or solution samples
£12 per cycle
 5 cycles minimum - no set up charge

For more information please contact Dr. Joe Gray:
 Phone/Fax: +44 191 222 8612
 Email: JOE.GRAY@newcastle.ac.uk

DNA sequencing (ABI 377-96XL, 877 Catalyst)
also available at competitive prices

READER ENQUIRY NO. 2